Genomics: a Swiss army knife to fight leprosy

THÈSE N^{O} 8482 (2018)

PRÉSENTÉE LE 6 AVRIL 2018

À LA FACULTÉ DES SCIENCES DE LA VIE
UNITÉ DU PROF. COLE
PROGRAMME DOCTORAL EN APPROCHES MOLÉCULAIRES DU VIVANT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Charlotte AVANZI

acceptée sur proposition du jury:

Prof. M. Blokesch, présidente du jury
Prof. S. Cole, directeur de thèse
Prof. S. Gagneux, rapporteur
Dr M. Moraes, rapporteur
Prof. J. Fellay, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Contents

Abbreviations

BB	borderline leprosy
BI	bacillary index
BL	borderline lepromatous leprosy
BT	borderline tuberculoid leprosy
CFP	culture filtrate protein
DDS	dapsone
DLL	diffuse lepromatous leprosy
DRDR	drug-resistance determining regions
ENL	*erythema nosodum leprosum*
ESAT	secretory antigenic target
FF	Fite Faraco
G2D	grade 2 disabilities
IGRA	interferon gamma release assay
IL-1ß	interleukin 1ß
IND	indeterminate leprosy
iNOS	inducible -nitric oxide synthase (iNOS)
LID-1	leprosy IDRI diagnostic-1
LL	lepromatous leprosy
MCP-1	monocyte chemotactic protein-1
MFP	mouse footpad
MDT	multi-drug therapy
PABA	p-aminobenzoic acid
PAS	para-aminosalicylate
PCR	polymerase chain reaction
PNL	pure neural leprosy
qPCR	quantitative polymerase chain reaction
RR	reversal reaction
SSS	slit skin smear
TT	tuberculoid leprosy
USA	the United States of America
WHO	World Health Organization
ZN	Ziehl Neelsen

Abstract

Leprosy, a highly disabling and stigmatizing infectious disease, is caused by *Mycobacterium leprae* and the newly discovered agent, *Mycobacterium lepromatosis*. Though treatable with antibiotics, leprosy has still not been eradicated, and around 200,000 new cases are reported every year worldwide, mainly in India, Brazil, and Indonesia. Tipping the balance towards leprosy elimination begins with improving our understanding of the pathogenesis and the transmission of the disease, which remains poorly understood. Current research on leprosy is markedly hindered by our incapacity to cultivate the leprosy bacilli on artificial media, as well as by the variation of the clinical forms of the disease. The rise of genomics in the 2000s has helped to get around these problems by opening new ways of studying organisms. Tools were developed to recover enough genetic material for downstream genomic applications; however, none of them is yet suitable for high throughput purposes.

In this thesis, we describe an optimized DNA extraction method from skin tissue that allows direct whole-genome sequencing, and enabled us to obtain ~ 250 genome sequences of *M. leprae* from different geographical locations throughout the world. Firstly, this dataset deepened our insight into the phylogeny of *M. leprae*, and points to the ancestral strain originating in East Asia and/or Europe. In addition, analysis of more than twenty drug-resistant strains revealed mutations in candidate genes potentially associated with new biological mechanisms such as drug resistance. Moreover, we analysed isolates from restricted geographic areas and from recurrent cases, and show that the distinction between relapse and reinfection with a closely related strain can be made but this remains challenging.

The whole genome sequencing of *M. lepromatosis* was achieved in 2015, and the discovery and use of new specific molecular detection methods allowed us to identify *M. lepromatosis* in the red squirrel population in the British Isles. In parallel, *M. leprae* was also discovered in red squirrels on Brownsea Island in the south of England. Though the risk of transmission from animals to humans is not yet clear, the discovery of a new animal reservoir for leprosy bacilli in a non-endemic country raises the question about the existence of other such reservoirs, especially in endemic countries, which could contribute to ongoing transmission.

Reliable and sensitive methods for detection of leprosy bacilli are crucial for early diagnosis and monitoring the disease. We show that efficient cell lysis during extraction increases the yield of genetic material recovered from leprosy bacilli and significantly improves the sensitivity of diagnosis by PCR for all leprosy forms.

Overall, our results highlight the impact and efficiency of genomics and whole genome sequencing for uncovering new biological mechanisms in unculturable bacteria such as the leprosy bacilli. Our results generated new hypotheses that await testing, and underline the massive potential of omics and bioinformatics for better understanding and fighting the disease.

Keywords: leprosy, *Mycobacterium leprae*, *Mycobacterium lepromatosis*, whole genome sequencing, drug resistance, animal reservoir, and transmission

Résumé

La lèpre est une maladie infectieuse progressive, invalidante et toujours stigmatisée, causée par *Mycobacterium leprae* et dans une moindre mesure par *Mycobacterium lepromatosis*. Bien qu'elle soit curable par une association d'antibiotiques, la lèpre n'est toujours pas éradiquée, et environ 200 000 nouveaux cas sont rapportés chaque année à l'échelle mondiale, touchant principalement l'Inde, le Brésil et l'Indonésie. Un des prérequis pour atteindre une élimination de la maladie passe par la compréhension de sa pathogenèse et de ses mécanismes de transmission qui demeurent aujourd'hui encore très mal décrits. La recherche fondamentale sur la lèpre est largement freinée par l'impossibilité de cultiver les bacilles sur des milieux artificiels ainsi que par l'existence d'un large panel de formes de la maladie suggérant des mécanismes biologiques différents. L'apparition des outils de génomique dans les années 2000 a permis de contourner ce problème et ouvert de nouveaux axes de recherche sur la lèpre. Plusieurs outils ont été développés permettant d'obtenir suffisamment de matériel génétique pour les tests génomiques. Cependant, aucun d'entre eux n'est adapté aux applications de haut débit.

Dans ce document, nous décrivons une méthode d'extraction d'ADN directement à partir de tissue cutané permettant le séquençage direct du génome bactérien. A ce jour, plus de 250 génomes de *M. leprae* ont été séquencés provenant de différents pays et couvrant tous les continents. A partir de ces résultats, nous avons tout d'abord affiné la phylogénie de *M. leprae* avec une origine hypothétique entre l'Asie de l'Est et l'Europe. Suite à cela, l'analyse de plus de vingt souches résistantes aux traitements nous a permis de mettre en évidence de possibles nouveaux mécanismes biologiques notamment associés à la résistance aux antibiotiques. Enfin, nous avons également analysés des isolats cliniques provenant de la même région ainsi que de patients présentant une récurrence de la maladie et démontré que la distinction entre rechute et réinfection avec une souche proche est difficile à déterminer.

Le séquençage du génome de *M. lepromatosis* a été achevé en 2015 et grâce à de nouveaux marqueurs spécifiques de la bactérie, nous avons pu identifier *M. lepromatosis* dans la population d'écureuils roux dans des îles britanniques. En parallèle, *M. leprae* a également été découvert dans le même réservoir sur l'île de Brownsea dans le sud de l'Angleterre. Bien que le risque de transmission de la maladie de l'animal à l'homme n'ai pas pu être établi avec les données disponibles, la découverte d'un nouveau réservoir pour la lèpre dans un pays non endémique soulève la question de l'existence d'autres réservoir, notamment dans les pays endémiques, contribuant à la transmission active de la maladie et notre incapacité à l'éliminer.

Enfin, le développement de méthodes de détection fiable et sensibles des bacilles de lèpre est crucial pour un diagnostic précoce de la maladie. Dans notre dernière étude, nous montrons que la réalisation d'une lyse efficace durant l'extraction augmente la quantité de matériel génétique de la bactérie disponible et améliore significativement la sensibilité de détection par PCR pour toutes les formes de lèpre.

En conclusion, nos résultats soulignent l'utilité et l'efficacité de la génomique et du séquençage haut débit pour identifier de nouveaux mécanismes biologiques chez des bactéries non cultivables pour lequel l'étude systématique est difficile comme les bacilles de la lèpre. Nos résultats offrent de nouvelles hypothèses biologiques et soulignes le potentiel des outils de omiques et de bioinformatiques pour l'amélioration des connaissances sur cette maladie et vers son élimination.

Mots clefs : lèpre, *Mycobacterium leprae*, *Mycobacterium lepromatosis*, séquençage haut débit, résistance aux antibiotiques, réservoir animal, et transmission

Chapter 1. Introduction

Leprosy, or Hansen's disease, is a slow and chronic infection leading to long-term physical disabilities, and often, social exclusion (1,2). Leprosy is the first disease to have been associated with an infectious agent, *Mycobacterium leprae* (1,3). Despite its long history of recognition, leprosy is one of the least understood infectious diseases afflicting humans and critical gaps remain in the knowledge about the spread and the exact mechanism of infection of the bacteria (4,5). The declaration of elimination of leprosy by the World Health Organization (WHO) in 2000, along with the overall difficulties in working with the leprosy bacilli have resulted in decreased interest and investment in basic research and led to negligence and oversight of the disease (6–9). In this introduction, I will present the current knowledge about the disease, spanning from the host to the bacteria, and emphasise the main gaps and challenges in the field. I will focus on the importance of genomics and the potential of this approach for the leprosy field.

CURRENT STATUS OF LEPROSY

Epidemiology: what do the numbers say?

Leprosy was highly prevalent in most parts of the world before the antibiotic era. After the introduction of the free of charge multidrug therapy (MDT) by the WHO in 1981, the prevalence of leprosy drastically decreased by 97% from 1985, with 171,948 cases reported in 2016 (4,10). In 2000, leprosy was considered officially eliminated at the global level (prevalence rate <1/10,000 inhabitants) (11). However, the disease still remained a public health problem, especially in India, Brazil, Madagascar, Mozambique, Nepal, and Tanzania, which account for 85% of the total burden (12). Thus, the "Final Push" campaign was launched in 2003 by WHO with emphasis on case detection and treatment in highly endemic countries (12). Even so, the decline of new leprosy cases has inexplicably slowed down or ceased entirely in some countries (4) (Fig. 1). Today, around 200,000 new cases are reported every year worldwide. In 2016, WHO reported an average of 8.5% of disease burden in children among new cases with a maximum of 38.1% in the Comoro Islands (10,13). These data suggested that uninterrupted transmission of the disease persists and leprosy cannot be contained by MDT alone.

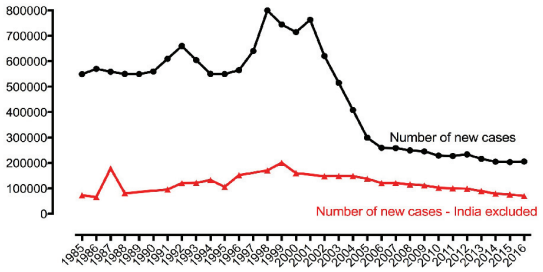


Figure 1: Number of new cases of leprosy reported between 1985 and 2016 worldwide (in **black) and excluding India (in red)** according to the weekly epidemiological records published by the World Health Organization every year.

Immunological classification and physio-pathology: the challenge for dermatologists

While leprosy affects many tissues such as bones, eyes, muscles and the genitalia, one of the first signs that lead leprosy patients to seek for medical consultation is dermatological changes (14). Different types and numbers of lesions such as plaques, papules, nodules and macules can be observed. The second most common symptom is the peripheral neuropathy leading to severe loss of sensitivity, thereto deformities, and disabilities in the absence of prompt treatment (15). The severity of the disease differs greatly between patients, resulting in a broad spectrum of clinical presentations whose outcome depends on the patient's immune status at the time of infection, and during the disease (1,16). One of the widely used classifications of leprosy is that of Ridley and Jopling (15–17), which is consists of five categories based on histological features and bacterial load: tuberculoid leprosy (TT), with the highest cellular response and a low bacterial load, followed by borderline tuberculoid (BT), borderline-borderline (BB), borderline lepromatous (BL), and the lepromatous type (LL), characterized by increased humoral immunity and a massive number of bacteria in the body (Fig. 2) (16,17). Indeterminate (IND) leprosy is not part of this spectrum. It is mainly observed in the early stages of the disease and can evolve either to spontaneous healing or progress towards the tuberculoid or lepromatous types

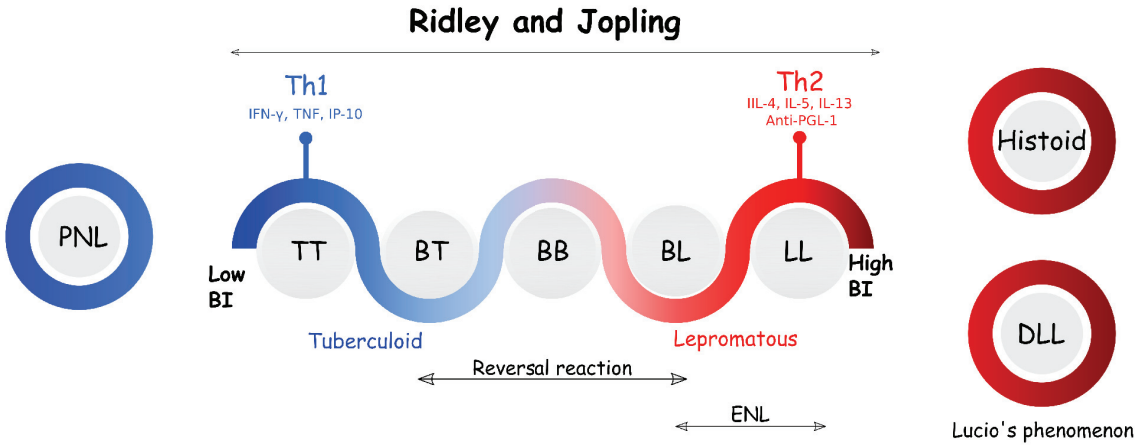


Figure 2: The Ridley and Jopling classification of leprosy and the unusual forms of leprosy, associated with the range of the bacillary load and the immune response. PNL: pure neural leprosy, DLL: diffuse lepromatous leprosy LL: lepromatous leprosy, BL: borderline lepromatous leprosy, BB: borderline leprosy, BT: borderline tuberculoid leprosy, TT: tuberculoid leprosy

(14). Additional specific and less frequent forms exist. Pure neural leprosy (PNL) is challenging to diagnose because patients harbour sensory neuropathy while skin lesions are mostly absent (18–20). PNL often progresses to a tuberculoid form (20). The diffuse lepromatous leprosy (DLL) form, also known as "pretty leprosy", is part of the LL spectrum. DLL is a severe form of leprosy associated with diffuse non-nodular bacterial infiltration in the skin, combined with vasculitis, severe alopecia and thrombosis (21). This form is mainly identified in Mexico and the Caribbean region while it is rarer in the rest of the world (22,23). Histoid leprosy is another form of lepromatous leprosy characterized by multiple nodules on the skin usually containing high bacillary load. These forms are considered extremely infectious (24).

Leprosy reactions are severe immunological complications of leprosy that often determine the course of the disease and the level of morbidity (15,25). Reactions are extremely painful and can occur any time before, during and after treatment (25). Reversal reaction is localised on the skin and nerves, and it is characterized by an increased cellular response (Th1) against *M. leprae* antigen. It occurs in around 30% of cases, mainly in the borderline spectrum, resulting in an upor down-grading of the disease toward the tuberculoid or the lepromatous side (15,25,26). *Erythema nodosum leprosum* (ENL) is observed in 50% of lepromatous and 5-10% of borderline cases when the amount of antigen-antibody immune complex is high (25), however, the exact immunological mechanism is not entirely understood (27). ENL is associated with painful inflammation extending over several years (28). An untreated reaction can lead to severe morbidity and mortality (28,29). The third response, named Lucio's phenomenon, is rare and exclusively associated with DLL. It leads to extremely severe symptoms of skin lesions surrounded by erythema progressing to necrosis of the skin. Untreated cases are fatal due to development of sepsis (25,30–33).

Generally, occurrence of skin lesions associated with sensory loss is a robust marker of leprosy. However, taken separately, the wide range of clinical symptoms might mislead patients and clinicians. Skin lesions alone are unspecific to the disease, and can easily be confused with another dermatitis, like vitiligo, psoriasis, leishmaniasis, cutaneous erythematous lupus, syphilis or lobomycosis (14). Neuritis can also be confused with other diseases such as diabetes for example (34). Also, the incubation time of the disease can range from a few months to 20 years (35–38), and the apperance of symptoms can be slow and insidious, which can close the opportunity for early treatment (39)

Leprosy lesions and tissue damage can occur in other organs, directly or indirectly due to *M. leprae* infection and peripheral neuropathy (40,41). For example, *M. leprae* infection can directly cause severe osteomyelitis of the nasal cavity and the limb

bones, leading to severe deformities. Osteoporosis can also occur, albeit more slowly and mainly in the limbs (39). Peripheral neuropathy indirectly results in sensory loss leading to ulcerations and secondary infections. Nerve injuries are also responsible for motor neuropathy with muscle atrophy and eyes involvement (39). Therefore, early diagnosis of leprosy is essential to stop disease progression, avoid irreversible lesions and reduce transmission.

Infection mechanism and host response

Characteristics of M. leprae

M. leprae was discovered in a human skin biopsy by Armauer Hansen during in 1873, and was the first microorganism associated with an infectious disease (3). Like other mycobacteria, *M. leprae* is a gram-positive bacterium composed of a thick cell wall with a layer of peptidoglycan linked to mycolic acid by arabinan and galactan chains. The cell wall is surrounded by a lipid layer mainly composed of phenolic-glycolipids (PGLs) (40). PGL-1 is the dominant lipid in the membrane conferring the immunological specificity. Altogether, these layers contribute to the resistance to acido-alcoholo treatments. *M. leprae* is an obligate intracellular pathogen and among the slowest growing bacteria with a doubling time of 10-12 days (twelve times longer than *M. tuberculosis*, the causative agent of tuberculosis) (15,42). *M. leprae* can grow between 10 to 32°C, which may explain its tropism for cooler parts of the human body (42). Despite many attempts, *M. leprae* is one of the few known pathogenic bacteria that cannot be successfully cultured in an artificial medium, significantly impairing the study of pathogenesis (1,43).

While *M. leprae* can be maintained in axenic culture for a limited period of time, large quantities of bacilli can only be obtained from *in vivo* models. Many important discoveries in the leprosy field were achieved using the mouse footpad assay (MFP) and the nine-banded armadillo (44–47). However, experiments relying on these methods require 9-18 months due to the slow-growing nature of the bacterium, and considerable resources. Widely used to test the efficiency of drugs, the usefulness of the MFP assay is limited in physiopathological studies because the model does not mimic leprosy symptoms and the infection remains localized in the footpad (40). On the other hand, the nine-banded armadillo displays the specific granulomatous response and the whole spectrum of leprosy forms, so it is extensively used for this purpose (48). Also, armadillos can successfully be challenged and immunized with heat-inactivated *M. leprae* bacilli or BCG vaccine (40). However, only one facility worldwide is working with this model at the moment, and the time required for such experiments is a major obstacle for using armadillos for vaccine development. Nevertheless, in contrast to the MFP model, infection of the peripheral nerve is widely described in armadillos, which is a unique model for nerve involvement in human (40).

Finally, the leprosy bacillus has a tropism for cells of the reticuloendothelial system, and a unique property to infect Schwann cells of the peripheral nervous system (15). Known mechanisms and the host response are presented in the next paragraph.

Host response and genetic markers

When entering the human body (probably through the nose or the skin); *M. leprae* is recognized by phagocytic cells such as macrophages and dendritic cells through the interaction between the pathogen-associated molecular patterns and the pattern recognition receptor. Once inside the host cell, the mechanism by which the host will develop a cytotoxic-like or antibody-like response is unknown, but it seems to depend on the immune response's cytokine secretion, the type of cells at the site of recognition, and the host genetic profile (discussed below) (49–51). From here, the bacterial growth depends on a cascade of cytokines that will activate the naïve T-cell CD4, either through the Th1 cytotoxic response leading to BT/TT lesions, or the Th2 helper response leading to the humoral response resulting in BL/LL lesions (40,51). Thus, depending on the host, the severity of infection might range from no infection to an extreme case of leprosy. In the cell-mediated immune response case, the disease might progress to spontaneous healing. However, it might also lead to sub-clinical and undetectable infections, where multiplication of bacilli is confined within localized granulomas, and the chronic inflammatory response can lead to long-term

disabilities (40). On the other side of the spectrum, in lepromatous lesions, the disease is not contained, and the number of bacilli is high. Inside the host cell, the bacilli stimulate host cell lipid metabolism leading to intracellular survival. Cells form foamy macrophages called Virchow's cells (49). In the last ten years, additional key players were identified in the pathogenesis mechanism of leprosy, such as Th17, Th22, and regulatory lymphocytes, opening a new area for research and improving our understanding of the disease (50,52).

Several host genetic markers have been associated with an increased protection or susceptibility to leprosy. The most convincing marker was discovered in a study including more than 1000 individuals from Vietnam and Brazil (53). A total of nine polymorphic alleles in the regulatory regions of PARK2, coding for an E3-ubiquitin ligase designated Parkin, and the co-regulated gene PACRG might increase susceptibility to leprosy in both populations (53). Mutations in TLR genes are also associated with susceptibility to different infectious diseases (54). In leprosy, specific mutations in the TLR1 in human and armadillo was associated with protective phenotypes whereas a specific mutation in TLR2 is linked to increased susceptibility to reversal reactions (40,51,55). Other SNPs in HLA, vitamin D receptor, TNF-α, IL-10, NRAMP1, and IFN-γ also defined genetic markers for the disease or reactions (51,52,56). While specific genetic host markers were associated with particular leprosy forms or level of protection against the disease, no association of specific *M. leprae* strains with clinical forms of the disease has been identified yet (57).

M. leprae has a tropism for infecting Schwann cells, which are glial cells conferring protection to peripheral nerves. It has been proposed that *M. leprae* binds to the α2-laminin of the Schwann cells through PGL-1, and with lesser affinity to ML-LBP21, a 21-kDa surface protein in *M. leprae*. Once internalized in Schwann cells, *M. leprae* would persist and multiply, and in response, the Schwann cells would express HLA-class II and activated CD4 T-cells which would initiate a chronic inflammatory reaction (40).

However, this hypothesis was challenged several times by the observations that the *M. leprae* bacillary load is low in peripheral nerves in paucibacillary patients despite nerve injuries being of the same intensity as in MB patients (58). Recently, a macrophage-induced mechanism was proposed by Madigan *et al.* as an alternative to explain nerve damage in leprosy, with PGL-1 as the main virulence factor (59). Briefly, using a zebrafish infection model, they observed that after phagocytosis of *M. leprae* by resident macrophages, *M. leprae*'s PGL-1 triggered the secretion of CCL2 by the resident macrophages, probably through the STING pathway, which lead to recruitment of permissive monocytes and enabled the transfer of the bacteria from the macrophages to the monocytes (59–61). At this point, infected monocytes are disseminated through skin lesions or blood and might reach peripheral nerves (59). Once in the proximity of peripheral nerves, PGL-1 induced progression toward a neurotoxic macrophage population through the secretion of inducible-nitric oxide synthase (iNOS), causing mitochondrial swelling and myelin loss in axons (59). Although not yet validated, it is possible that in MB cases, the abundance of *M. leprae* leads to death of macrophages, and bacteria become free to infect Schwann cells.

Further studies are needed to fully elucidate the exact mechanism by which the bacilli reach the Schwann cells (40), especially the differences between the PB and MB responses. Understanding of this mechanism could help to identify new drug targets.

LEPROSY BACILLI AND GENOMICS

Because of our inability to cultivate *M. leprae in vitro*, and the delays required for *in vivo* growth, working with the leprosy bacilli is prohibitively challenging, and this is where genomics has revolutionized leprosy research. The rise of molecular biology in the 80's has helped to develop new tools to study physiology (62–67) and to detect *M. leprae* directly from animal and human tissues such as skin, nasal swab or blood (68–72).

The whole genome era

In 2001, the first *M. leprae* genome was sequenced from the TN strain, isolated from a patient from Tamil Nadu in India and passaged in mice (73). With a genome size of

3,268,210 bp and an average GC content of 57.8%, *M. leprae* is the smallest and the AT-richest genome of any known mycobacteria (74). Surprisingly, only half of the genome is occupied by coding genes (1614), including 165 without a counterpart in *M. tuberculosis*, which could be good biomarker candidates for the development of specific diagnostic tools (74). The leprosy bacillus shows the highest pseudogene content of all sequenced genomes, with a total number of 1293 pseudogenes making up nearly half the total genome (75). The analysis of the original functional categories of pseudogenes revealed a massive loss of genes from the catabolism machinery, such as lipid uptake or energy metabolism, which might explain *M. leprae*'s slow growth and our inability to grow it *in vitro* (73). By comparing non-synonymous substitutions in the pseudogenes with functional orthologs from closely related genomes, it was estimated that *M. leprae* and *M. tuberculosis* diverged 66 million years ago, with a single pseudogenization event occurring 20 million years ago (75). This extreme genome decay is explained by a probable modification of the lifestyle of the bacteria, with adaptation to a new host or a strict intracellular niche (73–75). The exact genomic trigger of *M. leprae*'s reductive evolution is unknown, but the loss of sigma factors, the two-component systems and the DnaQ-mediated proofreading activity of DNA polymerase III probably contributed to the accumulation of pseudogenes (74). In addition to pseudogenes, 2% of the genome is occupied by 26 insertion elements and four families of dispersed repeats, with RLEP being the most abundant and accounting for 37 copies (76). Former transposable elements also participated in chromosomal rearrangements and genome downsizing (74,76).

In 2015, the whole-genome sequence of the pathogenic species *M. haemophilum* revealed a genetic closeness with *M. leprae* (77) (Fig. 3). This aspect is further discussed in the conclusion of this work.

Genotyping methods: VNTRs, SNP-typing or whole-genome typing?

In the absence of a bacterial phenotype, genetic markers can help to understand the transmission and epidemiology of leprosy. Molecular typing systems for leprosy epidemiological studies are based on different genetic markers within the genome of *M. leprae*.

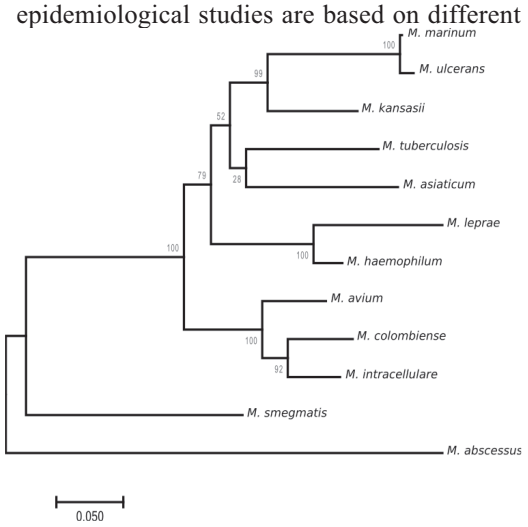


Figure 3: Maximum likelihood tree of *M. leprae* and selected mycobacterial species. The tree was created using MEGA7 from concatenated amino acid sequences of 11 proteins (DnaN, RpII, GrpE, MetG, RplY, PheT, FtsQ, HolA, MiaA, FtsY and, FtsX) [78,79]

Variable-number tandem repeats (VNTR) are widely used in epidemiological studies. They have been found to vary in copy number between strains of *M. leprae*. A total of 44 loci were described based on genome mapping, but only nine were reported stable and to be gave reproducible results (78). Besides, studies in India and Benin have shown variation in VNTR copy number in samples from the same patient, questioning the reliability of such markers (79,80). On the other hand, other studies have shown that groups of different VNTRs can be used for strain differentiation in a specific geographical area (81–83).

Comparative genomics of four different strains from India, Brazil, Thailand and the United States revealed remarkable genome conservation with 99.995% identity uncovered in only 215 polymorphic sites, mainly single nucleotide polymorphisms (SNPs) (84,85). Mapping these polymorphisms in more than 400 strains from 28 countries defined 84 informative markers (78 SNPs and six InDels in homopolymeric tracts) used for the classification in 16 SNP-subtypes of *M. leprae* (85), 1A-D, 2E-H, 3I-M and 4N-P (Fig. 4). SNP typing is a robust method, and the SNP subtypes were

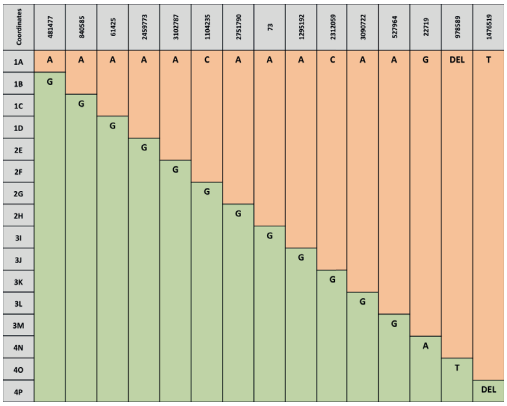


Figure 4: SNP typing system for the 16 SNP-subtype based on 16 markers (SNPs and indels) from Monot *et al.* (85)

associated with geographical localization (85). This is of high interest for global dissemination studies, but the resolution for short-range transmission, such as family cases, is limited. One solution could be to combine both VNTR and SNP typing (74).

Recently, whole-genome sequencing was successfully applied on *M. leprae* DNA extracted from ancient human remains (teeth and bones) and skin biopsies from patients from different countries (86). The principal challenge to deeply sequence *M. leprae* genomes is to recover enough bacterial DNA from the patient's sample and to remove enough host DNA, especially for samples with a low number of bacilli. To do so, Schuenemann et al. have developed an enrichment method based on DNA arrays coated with probes spanning the entire genome of *M. leprae* to capture DNA prior to sequencing (86). Thanks to this method, it is now possible to sequence genomes directly from human skin samples even if the amount of bacilli is low in the biopsy. Due to its high resolution, whole-genome sequencing is the ultimate tool that can be used to study the transmission of strains in a small geographical area, or to compare samples from different countries and to perform species-level phylogeny. However, the main drawbacks of this method are the cost, the time required from DNA extraction to sequencing, and the bioinformatics expertise. This is why such methods are not widely used in the leprosy field yet.

CHALLENGES OF LEPROSY

Stigma and leprosy: the unfortunate old partners

Leprosy stigma is still present in several parts of the world. Stigma arose most probably from the fear and revulsion certain individuals displayed toward persons affected by leprosy, notably the intensity of symptoms and body deformations associated with the disease (87).

The discrimination against people affected by leprosy was already described during the Middle Ages from where records of massive execution of people infected with leprosy were reported (88). At this time, leprosy was associated with God's punishment of the sinful and patients with the disease were quarantined in houses called leprosarium or lazar houses, often isolated from the general community i.e. islands (2,89). These "houses" were highly stigmatized, and most of the time, run by a religious community. Leprosy was considered for a long time hereditary until the discovery of the pathogen by Gerhard Hansen (90). Leprosaria slowly disappeared in Europe after the end of the Middle Ages and today, only one hospital is still active, in Fontilles in Spain where 60 patients were under treatment in 2005 (91). In other countries such as Vietnam, Laos, or India, villages of people suffering from leprosy still exist but are often taboo.

Nowadays, this discrimination has remains and has a major impact on the patient life at different levels: physically, psychologically, socially and economically. More worrisome, studies have shown a negative effect of stigma on disease management and especially for early diagnosis because patients prefer to conceal their illness to avoid discrimination (88,92). Unfortunately, such discrimination is rooted in the community as exemplified by the "Leper act" from 1893 applied in India. By this law, leprosy can be a cause for divorce, it can bar leprosy sufferers from voting, increase social charges, forbid public transport and driving, and even discriminate against those who resort to begging for support. Other studies have shown that discrimination does not only alter patients' llives impacts, but also the patients' family and other members of the community who support them, such as caregivers (93).

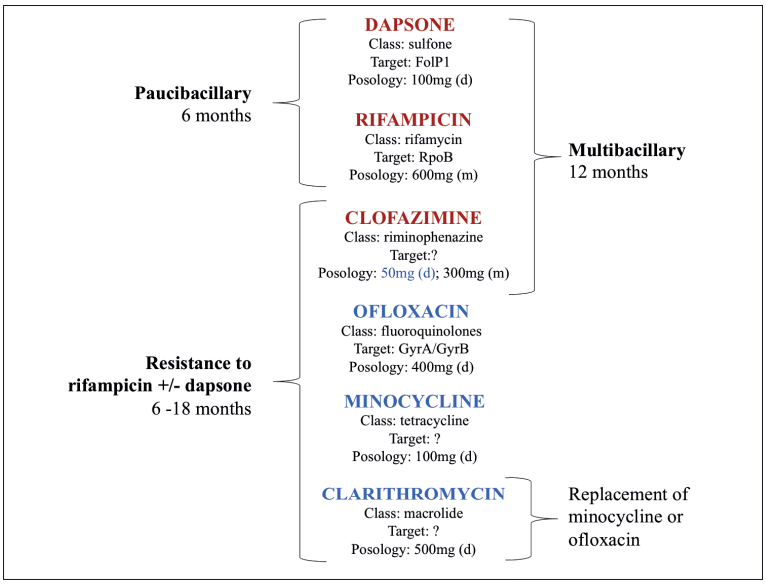


Figure 5: Multidrug therapy regimen for multi-and paucibacillary patients and in case of drug resistance. First-line drugs are in red, second-line drugs are in blue. d = daily, m = monthly.

Leprosy stigma affects women even more. Time to diagnose leprosy in women can be twice as long compared to men (94–96).

Most of the current stigma is associated with a lack of information about the disease. Thus, several organizations are working on decreasing this stigma using various approaches such as information campaigns or interventions with the governments (92).

Altogether intervention to decrease the stigma attached to leprosy is part of the challenge to halt leprosy transmission by promoting early detection and improvement of life for all patients.

Leprosy treatment

The monotherapy era

Sulfones were the first effective drugs used against leprosy in the 1940s Promin was soon replaced by its derivative Dapsone, a bacteriostatic compound, which was widely used in monotherapy due to its high activity and fewer side effects compared to other drugs (97). Dapsone acts at the first step of the folate pathway (98). Other inhibitors of the folate pathway such as trimethoprim or para-aminosalicylate (PAS), acting later in the folate cycle, proved to be ineffective against *M. leprae* (99). In parallel, clofazimine (B663), another bacteriostatic drug, was successfully tested in leprosy patients in a shortened treatment compared to dapsone alone, even if resistance to the therapy appeared after one year of monotherapy (97).

The development of new anti-leprosy drugs has always been impaired due to the uncultivability of the leprosy bacilli. The test of new drugs was drastically improved by the development of the *in vivo* mouse footpad model (MFP) by Shepard in 1960 (44,45). Since that time, all major anti-leprosy drugs were first tested using Shepard's method. Several anti-tuberculosis drugs were tested against *M. leprae*, some of which were inefficient, such as isoniazid, pyrazinamide or ethambutol, whereas ethionamide (protionamide), rifampicin and streptomycin were highly efficient (97,100,101). Ethionamide was soon abandoned due to the rapid emergence of drug resistance from monotherapy (97). Rifampicin is the most effective anti-leprosy drug due to its strong bactericidal activity (102). The drug binds to the ß subunit of the RNA polymerase and prevents transcription in the bacteria.

Multidrug therapy

Soon after the implementation of dapsone monotherapy, resistance arose in relapse cases but also in primary leprosy cases by 1980's followed by rifampicin resistance as well

(9,103). MDT was implemented by WHO in 1981 to prevent and treat drug-resistant strains as well as shorten the leprosy treatment (103). For practical treatment purpose, the WHO has established a simplified classification of leprosy patients. Paucibacillary (PB) cases, showing less than five skin lesions, are treated for six months with a combination of rifampicin and dapsone (Fig. 5), while multibacillary (MB) cases, harbouring five or more skin lesions, are treated for a minimum of 12 months with a combination of rifampicin, dapsone and clofazimine (Fig. 5). MDT is highly efficient and free of charge for all patients (104). Several clinical trials using the same drugs, but with shortened treatments, or with different therapeutical schemes have been tested, but the efficiency was never higher than the current MDT (105).

Second line drugs and other treatments

Emergence of drug resistance is unavoidable even with an efficient MDT. Rifampicin is the cornerstone drug of the leprosy treatment due to its rapid bactericidal activity (40), so efficiency of MDT will be severely compromised in cases of rifampicin resistance This led to the implementation of second-line antibiotics such as ofloxacin, minocycline, and clarithromycin (Fig. 5). These are highly effective drugs, but their high cost restricts their use as first-line drugs (106). Besides antibiotics, leprosy neuritis is treated with corticosteroids such as prednisone while thalidomide, prednisone or clofazimine can be used for ENL.

Administrated at the early stage of the disease, MDT can prevent severe deformities and disabilities. However, in an advanced stage of the disease, lesions are irreversible and can only be treated with surgery. Such intervention does not usually help to recover sensory loss, but have a substantial impact on social rehabilitation, especially when lesions are located on the face (107).

Since the implementation of MDT, a limited number of trials was undertaken to test new and less toxic drugs for leprosy treatment. One of the main apparent reasons is the fastidious experiment required for the *in vivo* efficiency studies. However, the rapid increase of drug-resistant *M. tuberculosis* strains has forced the community to develop new effective antibiotics, which could in turn be tested for leprosy. Thanks to the available genome sequence, it is now possible to predict the efficiency of new drugs *in silico*. For example, bedaquiline and PBTZ169, currently being evaluated for tuberculosis, are potentially effective drugs against *M. leprae* because the genes encoding for both targets, *atpE*, and *dprE1*, are functional in *M. leprae*. Also, bedaquiline was already tested *in vivo* and showed a high bactericidal activity against *M. leprae* in the MFP model (9). Q203 should work as well against *M. leprae* as the target is present (QcrB) and the alternative cytochrome oxydase (CydBC) has been deleted during the genome downsizing (108,109).

On the other hand, the prodrugs pyrazinamide and delamanid are probably inefficient, because their corresponding activators are not produced in *M. leprae* (110–114). Also, the resistance to PA-824, a nitroimidazole, was confirmed by MFP assay and it is probably due to the absence of the nitroreductase enzyme Ddn in *M. leprae* (115).

Molecular mechanisms of drug resistance

Since antibiotics are the only weapons against leprosy, surveillance of emergence of drug resistance is crucial. The gold standard method for antimicrobial resistance testing is the MFP assay. However, this method is time-consuming (6 to 9 months), laborious, and it requires highly trained technicians (116). For these reasons, the method is not suitable for routine use in a surveillance program. Following whole-genome sequencing and relying on studies conducted on *M. tuberculosis* and other bacteria, mechanisms of resistance to three antileprosy drugs were identified, namely rifampicin, dapsone, and ofloxacin (64,117,118). Specific mutations in the drug resistance-determining region (DRDR) of the DNA-dependent RNA polymerase subunit-ß (*rpoB*) decrease affinity for rifampicin and lead to resistance (Table1). Dapsone is an analog of para-aminobenzoic acid (PABA) and inhibits the dihydropteroate synthetase (*folP1*), the first step of the folate pathway (98). Two residues of FolP1 are involved in the resistance to dapsone (Table 1). Resistance to ofloxacin occurs in genes encoding DNA gyrase, principally in the DRDR of *gyrA* and, less frequently, in the

DRDR of *gyrB* (119) (Table 1). *M. leprae* clinical isolates harbouring mutation in *gyrB* have never been identified while *in vitro* models showed that mutation in the DRDR of *gyrB* should confer resistance. Clofazimine has both anti-inflammatory and antimicrobial activity, but the exact mechanism of action is not fully elucidated. Clofazimine tends to accumulate in body tissues, probably due to its highly lipophilic characteristics. With its long half-life, and depending on the administration frequency, clofazimine can remain in the body for months after treatment (120). There are different hypotheses to clofazimine's mode of action, the two most accepted involve a modulation of the bacterial potassium level, or the formation of reactive oxygen species after activation of the drug (121,122).

Table 1: Known mutations within the DRDR of *rpoB*, *folP1*, *gyrA* and *gyrB* that confer resistance to *Mycobacterium leprae* – most frequent mutations are in bold

Gene	Amino acid substitution	References
rpoB	Gln438Val	(123,124)
	Asp441Tyr	
	Asp441Asn	
	His451Asp	
	His451Tyr	
	Ser456Leu	
	Ser456Met	
	Ser456Phe	
	Ser456Trp	
	Leu458Val	
	Leu458Pro	
folP1	**Thr53Ala**	(98,117,123,125)
	Thr53Arg	
	Thr53Ile	
	Pro55Arg	
	Pro55Leu	
	Pro55Ser	
gyrA	Gly89Cys	(123,124)
	Ala91Val	
gyrB	Asp464Asn	(126)
	Asn502Asp	
	Glu504Val	

Clarithromycin and minocycline are part of the macrolide and tetracycline families, respectively. Their mechanism of action has not been elucidated in *M. leprae* infections (40).

PCR-sequencing of amplified DRDR is the screening tool recommended by the WHO. The main drawback of this method is the need for Sanger sequencing, a service that is not easily available, or affordable, in many endemic countries. As an alternative, the GenoType LepraeDR® from Hain LifeScience, a qualitative DNA strip assay based on classic PCR and reverse hybridization (127), allows identification of *M. leprae* DNA as well as simultaneous detection of antibiotic resistance in the three known targets (*rpoB, folP1*, and *gyrA*) without Sanger sequencing. The absence of some rare mutations and the DRDR of *gyrB* is the only drawback of this test. But overall, these methods do not offer enough resolution for identification of new resistance mutations or resistance mechanisms.

The global surveillance of drug resistance network was launched in 2008 and included 18 countries for sample collection and 18 reference centres for the analysis (128). Thanks to molecular techniques such as PCR-sequencing, the results of drug resistance screens can be obtained within hours. First used only in the assessment of relapse cases, antimicrobial resistance surveillance is now applied for all new leprosy cases because of the concerns from the increasing emergence and transmission of strains resistant to rifampicin, dapsone and ofloxacin (129). In the last WHO report, a total of 1600 strains were analysed between 2006 and 2013, of which 4% harboured mutations conferring resistance to rifampicin, 4% to dapsone and 1.5% to ofloxacin (129). While these numbers do not pose a direct threat for the global control of leprosy (128), of a bigger concern are the mounting reports of multidrug resistance (MDR) and primary cases of resistant strain transmission in *M. leprae* from different parts of the world (118,124,130–135). MDR is particularly problematic because the number of antibiotics available is low and three out of the six anti-leprosy drugs have an unknown mechanism of action, impeding a genomic-based survey of the resistance Therefore, monitoring the emergence and transmission of drug and multidrug resistance is an essential aspect of the global control of leprosy.

Vaccines

There is currently no effective vaccine for leprosy. The gap in our understanding of the

full pathogenesis of the bacteria and the host response at the different stages of infection including sub-clinical infections is holding up any prospects of efficiently developing one (40).

Several studies reported different ranges of protection by the *M. bovis* bacille Calmette et Guérin (BCG) vaccine, ranging from 30-50% (136). Another approach was to combine BCG vaccine with heat-inactivated *M. leprae* bacteria, however no significant benefit compared to BCG alone was observed. Also, the mass implementation of this vaciine would have been challenging since bacteria can be obtained only from armadillo lesions (136,137). Analysing the *M. leprae* genome revealed several candidate genes, and recombinant antigens were tested. The primary objective was to choose an antigen, or combination of antigens, that induce a strong INFγ response against the bacteria (138,139). One promising candidate is the 73f, a chimeric fusion protein composed of ML2028, ML2346 and ML2044 formulated with a TLR4L-adguvant. This protein is recognized by MB patient (IgG) and stimulated INFγ production in PB patients. This candidate vaccine entered into clinical trial in 2017 and early results should be available in 2018 (136,139).

M. indicus pranii (*MIP*), was tried as alternative leprosy vaccine. The formulation contained heat-killed *MIP* and was developed in the 90's. The vaccine protection was shown to be highest in children and protected 60% of contacts for six to seven years (136). Also, several studies reported the beneficial association of standard MDT with the vaccine leading to significant and faster recovery (140). In May 2017, a large trial was launched in India to evaluate the immune-therapeutic effect of single dose of rifampicin with the vaccine in patients and contacts. Data are not available yet.

Origin and spread of the leprosy bacillus

The dynamics of leprosy transmission throughout human history is not fully resolved, and characterization of the most ancestral strains is crucial for deciphering the nature of leprosy's origin. In this part, we will describe the current knowledge about the origin and evolution of *M. leprae* strains.

M. leprae is a clonal organism, and is genetically very conserved. The geographical distribution of *M. leprae* genotypes has helped retrace evolution and dissemination of leprosy. Earlier phylogeographical data showed a correlation of *M. leprae* genotypes with large human migrations, and support the hypothesis that leprosy originated either in East Africa (SNP type 2) or in India (SNP type 1), from where it spread to Europe and Asia via trade routes (84). In a newer study, Monot *et al.* suggested that the *M. leprae* ancestor was probably an intermediate between the SNP type 2 and SNP type 3 (85). Based on current genotyping studies, SNP type 2 is mainly found in Turkey (2F), Iran (2F), Ethiopia (2H), India (2E, 2H), Nepal (2G) and Europe (2F) (85,141,142).

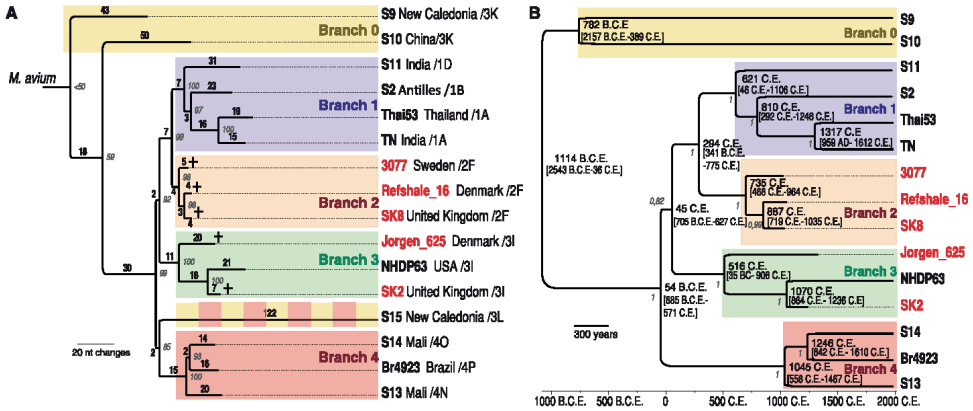


Figure 6: Phylogeny of medieval and modern *M. leprae* strain as described by Schueneman *et al.* [87] (A) Maximum parsimony tree of *M. leprae* genomes, rooted with *M. avium*. Geographic origin and SNP-subtype are given at each branch. Bootstrap (500) node support is shown in grey, and nucleotide substitution in bold. (B) Bayesian phylogenetic tree calculated with BEAST 1.7.1 inferred from 516 genome-wide variable positions. Posterior probabilities are in grey. B.C.E: before the Common Era; C.E: Common Era

Palaeoepidemiology is an emerging field that can provide clues to solve and understand the nature of infectious diseases and give a more comprehensive picture of the emergence, evolution, and spread of bacterial pathogens throughout history (86,142–147). However, studies on ancient DNA are technically complex, mainly because of the low amounts and poor quality of the DNA, and because the technique relies on cutting-edge protocols and requires specialized laboratories.

In 2013, using the array-based enrichment method, the first ancient *M. leprae* whole genomes were sequenced from Northern European cases. These strains correspond to SNP type 2F and 3I in the UK and Denmark (86,148) revealing a link between European strains and those currently reported from the Middle East (SNP type 2) and USA (SNP subtype 3I). Unfortunately, at that time no modern SNP type 2 genome was available. In the study of Schuenemann *et al.*, comparison of modern and ancient genomes also revealed that the 3K sub-type forms a distinct branch in the *M. leprae* phylogeny, called branch 0 (86), which was estimated to have diverged from the most common ancestor around the first millennium B.C. (Fig. 6). The 3K genotype has mainly been found in modern samples from China, Japan, and the Middle East (85,149) as well as in ancient isolates from Europe (142). It is noteworthy that the SNP subtypes 3M and 3J have not been sequenced yet, so their ancestral relationship with the other branches is not known (85). The only reported occurrences of these genotypes were in the French West Indies, and they are scarce in modern isolates while the genotype 3M was also identified in medieval European strains by PCR-sequencing (142).

The spread of leprosy throughout the world seems to be linked with human migrations, as several lines of evidence suggest (85). The most evident case is the introduction of leprosy into the Americas. The prevalence of SNP type 3I in contemporary South America, including the armadillo population currently spreading north, most likely derives from European settlers, as 3I was common in Medieval Europe. The abundance of SNP type 4 on the other hand is probably linked with the slave trade between South America and West Africa, where this genotype predominates (86) .

Human migrations and the contacts between different societies were dynamic during the last millennia, which blurs the genetic footprint of ancient leprosy transmissions. A global analysis of a large number of genomes might help to retrace the history of this disease.

Other leprosy-causing bacilli

The mysterious Mycobacterium lepromatosis: a new species or an M. leprae strain?

Leprosy had always been exclusively associated with *M. leprae* until 2008 when Han *et al.* discovered *M. lepromatosis* (150), in a liver sample from a Mexican leprosy patient. Despite several attempts, *M. lepromatosis* remains uncultivable *in vitro* (150).

Sequencing of more than 20 genes, including *rrs* (16S rRNA), *rpoB, groEL*, and *rpoT*, revealed that the overall level of nucleotide identity between *M. lepromatosis* and *M. leprae* is 90.9%, which is different enough to classify it as a new species (150,151). *M. lepromatosis* is the closest species to *M. leprae* known to date.. Strikingly, the bacterium has been reported mainly in Mexico and the Caribbean region, and it seems to be associated with the DLL form and especially with Lucio's phenomenon (152,153), but given the small sample size, the extent of this association or the possibility that *M. lepromatosis* causes other forms of leprosy remain unclear. (153).

Later, Han and colleagues identified *M. lepromatosis* in Singapore and Canada using a PCR-based screening, but most of the cases were still recorded in Mexico [116–118].

Whole-genome sequencing of *M. lepromatosis* was achieved in 2015. The description of the genome and the development of new tools to study the spread of the leprosy-causing bacteria is presented in **Chapter II**.

M. lepraemurium: *leprosy or non-leprosy bacillus?*

M. lepraemurium was described in 1903 and was associated with leprosy-like lesions in wild rats and mice in Russia, Germany, Japan and Puerto Rico (157). For a long time, *M. lepraemurium* was referred to as a “leprosy

bacillus" because of the clinical similarities between the human and rat symptoms and the difficulty to grow it *in vitro* (158,159). Moreover, several similarities were observed by electron microscopy between both pathogens. *M. lepraemurium* was successfully cultivated in rats and mice before the MFP method was developed for *M. leprae* by Shepard (159). Thus, *M. lepraemurium* infected mice was the only *in vivo* model of leprosy available at that time. Recently, *M. lepraemurium* were also associated with feline leprosy (158). Whole-genome sequencing of *M. lepraemurium* was achieved in 2017. The description of the genome and the relationships with *M. leprae* and other mycobacteria is presented in **Chapter II**.

Transmission

The mystery and challenges around the disease transmission

The exact mode of transmission of leprosy is unknown, but it seems that close contact with infected individuals in the family or the same community is a high-risk factor of transmission (160). Nasal mucosae or skin might serve as an entry route into the human host because leprosy bacilli were found in the nasal secretion and the skin of seemingly healthy individuals in endemic areas. (38). Generally, leprosy is considered to be poorly contagious and patients with a low bacillary index are usually less contagious compared to patients with a high bacillary load (161).

Most leprosy transmission studies focused on general clinical or societal markers to establish transmission patterns between patients, while some studies used genetic markers such as genotyping methods in an effort to increase specificity and resolution (38). However, monitoring the transmission of an *M. leprae* strain in a small geographic area and over a short period of time is challenging because of the scarcity of mutations that could distinguish between different isolates. SNP typing methods (not enough resolution) and VNTRs (variable mutation rates between loci and frequent reversal mutations) are not apt for this challenge. Whole genome-based approaches have the potential to reveal the full dynamic range of transmission of closely related strains.

An important area where such resolution is needed is the differentiation between relapse and reinfection cases. Relapse rate is an indicator of treatment efficiency directly link to drug-resistant or persistent strains, while re-infection is an indicator of active transmission in the area (81). Leprosy re-occurrence most often happens in cases with a high bacillary index and is mostly attributed to relapse, possibly due to the dogma about protection gained to the disease after the first infection (162) (communication Leprosy Mailing List – July 17, 2017). For example, a study conducted in India found three recurrent cases among 163 followed-up cases after two, four, and eleven years post MDT. All three recurrent cases were attributed as relapse without any molecular confirmation (163). Recently, using a combination of VNTRs and SNP-typing, Da Silva Rocha and colleagues identified two cases of reinfection that occurred nine and fifteen years post primary infection (81). Such findings are of particular importance for endemic areas where convalescents continue to be exposed to *M. leprae* (164). Leprosy recurrence is commonly observed in some parts of the world, but no study has systematically assessed the rate of relapse compared to reinfection in these areas (162). More information about leprosy recurrence is needed in order to improve treatment management, and to better assess the effectiveness of new short-course treatment regimens (165).

The natural non-human source of leprosy bacilli

Humans were thought to be the only source of leprosy infection (166) until non-human natural hosts of *M. leprae* were identified in the 1970's [75]. *M. leprae* is well established in the wild nine-banded armadillo (*Dasypus novemcinctus*) and to a lesser extent in some primates, mainly in Africa (47,169–171). Prevalence in the wild armadillo is around 20% but it depends on the geographical area (172). The territory covered by nine-banded armadillos includes Brazil, Central America and the Southern United States. Using genomics approaches, it was shown that leprosy in the USA can be a zoonotic infection between nine-banded armadillos and human (173). Leprosy was likely transmitted to the nine-banded

armadillo by European settlers. Leprosy disappeared from the USA during the modern period, but it re-emerged in the South about a century ago, which roughly corresponds with the time of introduction (≈1850) and expansion of the nine-banded armadillo population from Mexico into the Southern USA (172,174,175). As for humans, the mechanism of transmission of leprosy between nine-banded armadillos and humans is not known, but it is assumed that close contact with infected animals and contaminated flesh and blood is a high risk factor (172,173,176). The burden of leprosy in nine-banded armadillos in Brazil, a highly endemic country for human leprosy, and its possible impact to the human population, have not been studied yet.

Recently, the newly discovered leprosy agent *M. lepromatosis* has been identified in red squirrels in Scotland, presenting with leprosy like lesions (177,178). This is the first report of a non-human reservoir of *M. lepromatosis*. It is surprising to find *M. lepromatosis* in Scotland, a non-endemic country for leprosy. Additional work is needed to define the role of the red squirrel as a new source of leprosy bacilli and transmission.

Other potential natural reservoirs

The existence of animal reservoirs and newly detected human cases with no tangible direct source of infection known from leprosy patient has led to the idea that additional reservoirs of *M. leprae* might exist (179,180). Several studies have shown the presence of *M. leprae* in soil or water (38). In a survey conducted in India, the authors found the presence of *M. leprae* 16S rRNA in 35% of the soil samples and concluded that viable *M. leprae* were present in the soil at the time of collection (181,182).

In the environment, free-living amoebae can feed on bacteria. Encystment is a survival stage in amoebae, also known as a resting form. Several bacteria can survive within free-living amoebae, including some mycobacteria. Encysted amoebae provide protection to the internalized bacteria (183) and contribute to their spread in the environment. Interestingly, Lahiri *et al.* have shown that *M. leprae* can survive in a free-living amoebae such as *Acanthamoeba castellani* for a few days and can grow when re-inoculated in MFP (184). Following this study, Wheat *et al.* reported that *M. leprae* could survive up to eight months in encysted amoebae without loss of virulence (185), but so far, no *M. leprae* has been found in amoebae in the environment.

Overall, it is evident that the natural reservoir of *M. leprae* is broader than previously anticipated, and studying non-human sources more deeply might provide a leap for leprosy control.

Diagnosis of leprosy

Diagnosis of leprosy at its early stages is essential to prevent disabilities. Nowadays, leprosy diagnosis still relies on the identification of three cardinal signs, namely skin lesions, thickness or enlargement of peripheral nerves, and the detection of acid-fast bacilli in active lesions or slit skin smears (SSS). Accurate diagnosis of leprosy can only be done by trained physicians because the symptoms overlap with those of a number of other medical conditions. This emphasizes the importance to maintain leprosy expertise in endemic and non-endemic countries as well as to reinforce training for all health care workers . Another critical component of early diagnosis is informing and training patients, their contacts and the general population about the disease to decrease the associated stigma that often results in leprosy sufferers hiding their disease, as this will facilitate the identification of early symptoms and prompt intervention of medical staff (15,186,187).

Table 2: The bacteriological index scale used for expression of the bacterial load – the result is expressed in logarithmic scale of the number of bacilli observed in a x100 objective field (188)

BI	Observations
6+	Many clumps of bacilli in an average field (over 1000)
5+	100-1000 bacilli in an average field
4+	10-100 bacilli in an average field
3+	1-10 bacilli in an average field
2+	1-10 bacilli in 10 field
1+	1-10 bacilli in 100 field
0+	No bacilli in 100 fields

The amount of acid-fast bacilli in active lesions or SSS determines the bacterial load, also named bacillary or bacteriological index (BI) (Table 2). The BI system is more effective with high bacillary loads, but a BI of zero is not an indication of the absence of leprosy.

Since *M. leprae* and *M. lepromatosis* cannot be cultivated *in vitro*, their presence must be assessed directly in different types of tissues, or body fluids such as skin, nerves, urine, or blood. However, and especially in leprosy with a low BI, bacteria are commonly not detectable in SSS, urine or in blood (189–191). Also, nerve sampling is usually undertaken only for diagnosis of pure neural leprosy since the procedure is highly invasive. Skin biopsy is the most common type of sample used for diagnosis. From such samples, histology with hematoxylin and eosin staining is performed to establish the disease classification. Then, the presence of leprosy bacilli is verified using microscopy or genomic methods (Table 3).

Table 3: Non exhaustive list of methods used on SSS and skin biopsies for the detection of leprosy bacilli in skin tissues

	Aim	References*
Ziehl-Neelsen stain / Fite Faraco stain	Identification of acid-fast organisms – mycobacteria are colored in red	(192,193)
Haematoxylin and eosin stain	Coloration of cells in a histology section (nucleic and anucleic cells) – Classification of the leprosy spectrum form	(16,192)
Auramine O stain	Identification of acid-fast organisms – mycobacteria are fluorescent	(192–194)
PCR/qPCR	Identification of a species based on detection of specific genomic parts.	(194–197)

Ziehl Neelsen (ZN) and Fite Faraco (FF) staining and light microscopy are commonly used in countries with limited resources. These low-cost methods are specific, but lack sensitivity, especially for samples with low BI and require trained and specialized microscopists. In this regard, genomic methods are superior to microscopy. First, conventional polymerase chain reaction (PCR) and real-time fluorescence-based quantitative PCR (qPCR) methods are known to be extremely sensitive with low detection limits (198). For example, Truman and colleagues reported a limit of detection of only three bacilli using a qPCR method (195). Also, since the pieces of tissue used for DNA extraction are much larger than those used for microscopy, there is more chance to encounter bacilli. For example, PCR methods detected *M. leprae* DNA extracted from skin biopsy more efficiently than SSS, because there are fewer bacilli in SSS (196). Among the different targets suitable for amplification, such as the manganese-dependent superoxide dismutase gene (*sodA)*, the mRNA of antigen 85B (Ag85B), and the 16S rRNA gene; RLEP, an *M. leprae* specific dispersed repeat, is the most sensitive marker and a promising candidate for PCR-based diagnosis (197,199). The sensitivity of detection for RLEP varies from 70 to 100% depending on the method and type of sample (200). For example, qPCR is more sensitive than conventional PCR and easier to perform but requires more sophisticated equipment (200). The sensitivity of PCR-based methods is up to 100% for MB cases, and around 70-80% in PB cases (200). Most importantly, it is possible to detect the presence of *M. leprae* by PCR even in cases where the result of microscopic methods is negative (196).

While working with one of the most sensitive methods available, one could ask why molecular methods do not give a higher sensitive result for PB patients? Obviously, the number of bacilli present in the skin is one of the main reasons but another critical step, which is often not taken into consideration, is the DNA extraction method used since the quality and the amount of DNA recovered can be widely affected by the chosen method. Also, accurate definition of the gold standard differed depending on the publication, and it is an important point to compare the different results (200). Both parameters are further discussed **in the last chapter and the conclusion**.

The GenoType LepraeDR® from Hain LifeScience is the only commercially available and standardized molecular test for leprosy diagnosis. The test shows a specificity of 100% and sensitivity of 100% in the cohort

tested (127). However, the cohort consisted exclusively of smear-positive patients (BI >0), avoiding the most difficult cases to be tested.

Immunodiagnostic tools are also available for diagnosis of leprosy. The first one developed was the lepromin test, similar to the tuberculin test available for tuberculosis. Briefly, heat inactivated *M. leprae*, or synthesized *M. leprae* antigens, are intradermally injected, and the reaction is observed after two days (Fernandez reaction) and four weeks (Mitsuda reaction). Although the Fernandez reaction appears in both LL and TT cases, the Mitsuda reaction is usually observed in TT patients and reflects the cellular immunity. The sensitivity of this test is low, which is why it is now mainly used for classification purposes (201). Nevertheless, the Mitsuda test is still performed in several countries.

PGL-I antigen from *M. leprae* triggers the formation of IgM class antibodies. These anti-PGL-1 antibodies are found in high quantity in the sera of MB patients (>95% positivity) due to the increased humoral response, while they are often absent in PB patients (20-40% positivity) (202). The test was available commercially as a lateral flow test but the production was stopped before 2010 due to low demand and poor take-up by leprosy control programs (203). Similar to IgM, the IgG antibody response directed against *M. leprae*-specific recombinant proteins such as ML0405 and ML2331 (leprosy IDRI diagnostic-1 or LID-1) correlates with the bacterial load. Thus, like for anti-PGL-I, MB patients are the easiest to detect compared to PB patients using the IgG test. Recently, both tests were combined to enhance detection of MB cases in the so-called LID-ND-O antigen from the Infectious Disease Research Center (204). Nevertheless, IgM and IgG responses are weak or absent in PB patients, which present high cellular immunity. However, recently, Serrano-Coll and colleagues suggested in their study conducted on 80 patients that LID-ND-O positivity is a good risk marker of developing neuritis and leprosy reactions (204).

For PB side of the spectrum, the pro-inflammatory mediators IFN-Ɣ and IP-10 were tested. For tuberculosis, the commercial Quantiferon-TB gold diagnosis test is based on the interferon gamma release assay (IGRA) where IFN-Ɣ secretion is detected in whole blood after exposure to the specific 6KDa secreted antigenic target (ESAT-6) and 10kD culture filtrate protein (CFP-10) (205–207). This test is sensitive and specific for the detection of *M. tuberculosis* complex infection but does not differentiate active disease from latent tuberculosis (208). Studies have shown cross-reaction between *M. leprae* and the *M. tuberculosis* CFP-10 and ESAT-6 antigen [192–194]. The IGRA test does not detect MB cases as expected, and only a small part of PB patients. But only a few studies have been conducted with the Quantiferon so far, and none of them used CFP-10 and ESAT-6 purified from *M. leprae*, which could increase the sensitivity (208,209). However, one limitation, raised by Geluk and colleagues, is the strong cross-reactivity between ESAT-6 of *M. lerpae* and *M. tuberculosis*, which can increase the complexity of diagnosis (205). Prospects for other IFNƔ specific *M. leprae* antigen stimulators were conducted to differentiate between PB and MB patients and PB and healthy endemic control. So far, no such antigen was described (52). Thus, current research is focused on the detection of one or several cytokines/chemokines specific for one or several forms of the leprosy spectrum such as the monocyte chemotactic protein-1 (MCP-1) and the interleukin 1ß (IL-1ß). The primary aim is to provide a precise diagnosis between healthy controls and leprosy patients at the same time that an accurate classification of the disease (52). Recently, von Hooij *et al.*, combined the detection of a set of cytokines/chemokines (IL-10, IP-10, and CCL4) and anti-PGL-I in whole blood from leprosy patients from Bangladesh (210). The selection of the markers was driven by the aim to distinguish between different clinical leprosy types. ML2478 induces high IFN-γ and consequently IP-10 concentrations, and is associated with *M. leprae* exposure and risk of infection (210). The chemokine (C-C motif) ligand 4 (CCL4) secretion is increased in patients, and partly in household contacts, but not in endemic controls. IL-10 is involved in the suppression of Th1 cells in MB leprosy patients. Finally, anti-PGL-1 was included in the final set because it is specific to MB patients. Hence, the test can simultaneously perform multiple analyses in one sample and can distinguish different clinical leprosy types. Using a combination of these markers,

it is possible to identify an ongoing infection, and classify the disease (210). Although encouraging, these tests need to be validated in several endemic areas to assess the robustness of the method.

Interruption of leprosy transmission and diminution of leprosy burden is impaired by the absence of an efficient, reliable and global diagnosis tool for all clinical forms but also for household contacts. Currently, diagnosis is nearly impossible before the onset of clinical signs, and this is particularly challenging for the forms with a low number of subtle symptoms such as pure neural leprosy, unusual forms (in children for example) or sub-clinical infections (13,15,18). An efficient diagnosis tool allowing detection of the disease prior to the onset of symptoms is one of the main aim and wish objectives in the leprosy field (12)

Other approaches to detect leprosy include the "omic" techniques, such as metabolomics or transcriptomics. While these methods could help to identify global diagnostic or prognostic markers, they could also uncover pathogenic mechanisms and participate in a better understanding of leprosy pathogenesis.

Metabolomics was already applied for various infectious diseases to identify a specific metabolic signature for the diagnosis and prognosis (211). In leprosy, investigation of specific metabolomic signature has been investigated in the serum and the urine (212,213). Recently, Mayboroda and colleagues identified a set of metabolites in urine that was able to discriminate between endemic controls and leprosy patients. Also, they reported a modification of the metabolome during type 1 reaction (213). Although these results are encouraging, they need to be validated in a more demanding cohort and tested in another disease to demonstrate the specificity to leprosy.

An important breakthrough was recently made using transcriptomics to understand the immune response and pathogenesis of several infectious diseases (214,215). Genome-wide transcriptional profiling of individuals with active tuberculosis led to the identification of a specific 86-gene signature that can distinguish tuberculosis from other infectious or inflammatory diseases, and determine the severity of the disease from type I interferons IFN-αß levels (216). In leprosy, a similar approach was used to identify transcriptional changes of the immune reactions (217). Dupnik *et al.* showed immunity-related pathways to be up regulated in the transcriptional profiles of leprosy patients with reversal reaction or ENL, with complement pathway deposition being part of the immune response to both reactions. When the pathology is mediated by complement and other components of innate immunity, they hypothesized that such reactions result from abnormal antigen recognition. This observation helped explaining the pathogenesis of reactions. Such an approach would also be helpful to identify global biomarkers of leprosy, in blood for example.

Prevention in household contacts

Current studies showed that leprosy contacts and people living in endemic areas have more chance to contract and develop the disease (200). Exposure to *M. leprae* can be evaluated by serology and qPCR on DNA extracted from nasal swabs. Presence of bacteria in the nose is not a sign of disease, but individuals with nasal carriage have a higher risk to develop leprosy (200). However, the fraction of nasal carriers of *M. leprae* in the population varies from 1 to 70%, depending on the geographical region and the group considered (contacts or all healthy individuals for example) (200). More interestingly, Romero-Montaya and colleagues confirmed the familial transmission between contacts and index cases using a set of VNTR and SNP markers (218). However, SNP genotyping could be done only for one index case, probably because of the low amount of bacterial DNA in other cases. Contacts with positive PGL-1 also have a higher risk of developing the disease and especially if the index case has a high BI (190,200). Thus, Martinez *et al.,* suggested that combining conventional PCR or qPCR detection methods with anti-PGL-I serology could help identify individuals exposed to *M. leprae* and at high risk to develop the disease (200). This combination was applied in the study from Romero-Montaya *et al.,* but no follow up of the tested individuals was done to evaluate the sensitivity of this association (218). Another study is currently ongoing using both methods described above and preliminary results showed more than 65% positivity in

household contacts in a high endemic area in Parà, Brazil (13).

Another approach would be to try to identify a host biomarker based on comparative transcriptomics. Such an approach was successfully applied in tuberculosis. Zak and colleagues compared the transcriptome of individuals infected by *M. tuberculosis* (positive with the IGRA or tuberculin tests) who developed symptoms after two years with positive individuals who did not develop symptoms after two years (219). Using RNA sequencing in whole blood, they identified 48-biomarkers that predicted the risk of tuberculosis progression from latent to active with a sensitivity of 66% and a specificity of 80% (219,220). For leprosy, such an approach is more challenging since it would first require a well-established gold standard test to define "positive" individuals, such as the PGL-1 serology and PCR on DNA extracted from nasal samples described above. Moreover, since the incubation time of the disease be up to 20 years, the study would require an large number of individuals to be followed up during several years.

Monitoring therapy efficiency

The assessment of therapy efficiency is an important parameter that can be used to identify lack of observance, appearance of drug resistance, or potential emergence of persistent strains, as exemplified in *M. tuberculosis* (111,221,222). Currently, the primary tool to measure the viability of bacilli is through the assessment of RNA expression though RT-PCR (182). However, like for DNA, this approach is challenging in samples with low bacillary loads. In tuberculosis, Thompson and colleagues identified a five-gene host signature that allows assessing the risk of treatment failure (223). For leprosy, such a signature would also help to monitor treatment efficiency in all leprosy forms despite the bacterial load and to provide an efficient biomarker for future clinical trials.

OVERVIEW OF THE THESIS

In the absence of tools to study the phenotype of leprosy bacilli, genomics approaches and especially whole-genome sequencing can provide valuable information to understand the biology, genetics and the evolution of the leprosy bacilli at different time scales. Whole-genome analysis provides deeper resolution than standard molecular typing methods such as SNP-typing or VNTRs. However, high throughput sequencing methods are expensive and challenging with DNA extracts from leprosy skin biopsies, especially those with low BI. Efficient DNA extraction methods for the leprosy bacillus, tailored for various types of samples, are therefore primordial for routine whole-genome sequencing applications. Thus, during the first years of my doctorate programme, I was mainly working on the development of an efficient and less expensive method to recover enough *M. lerpae* DNA directly from skin biopsies. This method was published in 2016 (**Chapter IV** and **Chapter V**). Later, other methods have been optimized for different types of starting materials such as formalin fixed samples (**Chapter V**) or pure bacterial suspensions (**Chapter II**).

Thanks to the new extraction methods, the whole genomes of *M. lepromatosis* and *M. lepraemurium* were sequenced, providing more insight into the evolution and biology of these species. Results are presented in **Chapter II**.

In 2013, only 15 *M. leprae* genomes were available; scarcely covering different parts of the world. During my thesis, I concentrated efforts on sequencing additional strains from medieval Europe in order to identify ancestral lineages of *M. leprae* (**Chapter III**) and a modern samples from Japan, Ethiopia, West Africa and Brazil (**Chapter IV**) for phylogenetic and transmission studies as exemplified by the work in Guinea-Conakry.

In addition, I focused on the sequencing of known drug and multi-drug resistant strains in order to identify new mechanisms involved in drug resistance (**Chapter IV**). Moreover, recurrent cases were studied to understand the aetiology and to obtain information about strain variation within the host and within a limited endemic area (**Chapter IV**).

Animal reservoirs of the leprosy bacillus were investigated in the red squirrel population from the British Isles using different approaches such as histology, serology and genomics (**Chapter IV**).

Finally, the host depletion method for DNA extraction was proven to be highly

effective for high BI but it is still limited for BI 0 to 2. Thus, for diagnosis purposes, the DNA extraction method was modified to maximize the yield of *M. leprae* DNA recovered in the low BI samples. Combined with PCR, the sensitivity of the method was compared with other leprosy diagnostic methods in **Chapter V.**

REFERENCES

1. Britton WJ, Lockwood DN. Leprosy. The Lancet. 2004 Apr 10;363(9416):1209–19.

2. Sermrittirong S, Van Brakel WH. Stigma in leprosy: concepts, causes and determinants. Lepr Rev. 2014 Mar;85(1):36–47.

3. Hansen GA. Memoirs: The Bacillus of Leprosy. J Cell Sci. 1880 Jan 1;s2-20(77):92–102.

4. Schreuder PAM, Noto S, Richardus JH. Epidemiologic trends of leprosy for the 21st century. Clin Dermatol. 2016 Jan;34(1):24–31.

5. Fine PEM. Leprosy: the epidemiology of a slow bacterium. Epidemiol Rev. 1982 Jan 1;4(1):161–88.

6. Binns C, Low WY. Leprosy: Feared and Forgotten, but Still with Us. Asia Pac J Public Health. 2016 Oct 1;28(7):560–1.

7. WPRO | Remembering those with leprosy, a forgotten disease [Internet]. WPRO. [cited 2017 Nov 24]. Available from: http://www.wpro.who.int/laos/mediacentre/releases/2017/20170201-world-leprosy-day/en/

8. Keed A. Leprosy: a forgotten disease? Infect Dis Clin. 2017;18(5):61–6.

9. Gelber RH, Grosset J. The chemotherapy of leprosy: an interpretive history. Lepr Rev. 2012 Sep;83(3):221–40.

10. WHO. Global leprosy update, 2016: accelerating reduction of disease burden. 2017 p. 501–20. (35). Report No.: 92.

11. WHO. Leprosy: global situation. 2002 p. 1–8. Report No.: 77.

12. Rinaldi A. The Global Campaign to Eliminate Leprosy. PLoS Med. 2005 Dec;2(12):1222–5.

13. Barreto JG, Frade MAC, Bernardes Filho F, da Silva MB, Spencer JS, Salgado CG. Leprosy in Children. Curr Infect Dis Rep. 2017 Jun;19(6):23.

14. Talhari C, Talhari S, Penna GO. Clinical aspects of leprosy. Clin Dermatol. 2015 Jan 1;33(1):26–37.

15. Rodrigues LC, Lockwood DN. Leprosy now: epidemiology, progress, challenges, and research gaps. Lancet Infect Dis. 2011 Jun;11(6):464–70.

16. Ridley DS, Jopling WH. Classification of leprosy according to immunity. A five-group system. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1966 Sep;34(3):255–73.

17. Ridley DS. Histological classification and the immunological spectrum of leprosy. Bull World Health Organ. 1974;51(5):451–65.

18. Jardim MR, Antunes SLG, Santos AR, Nascimento OJM, Nery JAC, Sales AM, et al. Criteria for diagnosis of pure neural leprosy. J Neurol. 2003 Jul;250(7):806–9.

19. Tomaselli P, Frade M, Foss N, Nascimento O, Marques V, Barreira A, et al. Pure Neural Leprosy: Clinical and Electrophysiologic Features of Patients Evaluated in a Brazilian Tertiary Centre of Neuromuscular Diseases (S07.004). Neurology. 2012 Apr 24;78(1 Supplement):S07.004–S07.004.

20. Rao PN, Suneetha S. Pure neuritic leprosy: Current status and relevance. Indian J Dermatol Venereol Leprol. 2016 May 1;82(3):252.

21. Prem Anand P, Oommen N, Sunil S, Deepa MS, Potturu M. Pretty leprosy: Another face of Hansen's disease! A review. Egypt J Chest Dis Tuberc. 2014 Oct 1;63(4):1087–90.

22. Herath S, Navinan MR, Liyanage I, Rathnayaka N, Yudhishdran J, Fernando J, et al. Lucio's phenomenon, an uncommon occurrence among leprosy patients in Sri Lanka. BMC Res Notes. 2015 Nov 13;8(672).

23. Han XY, Quintanilla M. Diffuse Lepromatous Leprosy Due to *Mycobacterium lepromatosis* in Quintana Roo, Mexico. J Clin Microbiol. 2015 Nov 1;53(11):3695–8.

24. Karthikeyan K. Histoid Leprosy. Am J Trop Med Hyg. 2015 Jun 3;92(6):1085–6.

25. Kamath S, Vaccaro SA, Rea TH, Ochoa MT. Recognizing and managing the immunologic reactions in leprosy. J Am Acad Dermatol. 2014 Oct 1;71(4):795–803.

26. Walker SL, Lockwood DNJ. Leprosy type 1 (reversal) reactions and their management. Lepr Rev. 2008 Dec;79(4):372–86.

27. Negera E, Walker SL, Girma S, Doni SN, Tsegaye D, Lambert SM, et al. Clinico-pathological features of erythema nodosum leprosum: A case-control study at ALERT hospital, Ethiopia. PLoS Negl Trop Dis. 2017 Oct 13;11(10):e0006011.

28. Walker SL, Sales AM, Butlin CR, Shah M, Maghanoy A, Lambert SM, et al. A leprosy clinical severity scale for erythema nodosum leprosum: An international, multicentre validation study of the ENLIST ENL Severity Scale. PLoS Negl Trop Dis. 2017 Jul 3;11(7):e0005716.

29. Walker SL, Lebas E, Doni SN, Lockwood DNJ, Lambert SM. The mortality associated with erythema nodosum leprosum in Ethiopia: a retrospective hospital-based study. PLoS Negl Trop Dis. 2014 Mar;8(3):e2690.

30. Kumari R, Thappa DM, Basu D. A fatal case of Lucio phenomenon from India. Dermatol Online J. 2008;14(2):10.

31. Ang P, Tay Y-K, Ng S-K, Seow C-S. Fatal Lucio's phenomenon in 2 patients with previously undiagnosed leprosy. J Am Acad Dermatol. 2003 Jun;48(6):958–61.

32. Rea TH. Lucio's phenomenon: an overview. Lepr Rev. 1979 Jun;50(2):107–12.

33. Vera-Cabrera L, Escalante-Fuentes WG, Gomez-Flores M, Ocampo-Candiani J, Busso P, Singh P, et al. Case of Diffuse Lepromatous Leprosy Associated with "*Mycobacterium lepromatosis*." J Clin Microbiol. 2011 Dec 1;49(12):4366–8.

34. Kumar B. Pure or Primary neuritic Leprosy (PNL). Lepr Rev. 2016;87:450–5.

35. Brubaker ML, Meyers WM, Bourland J. Leprosy in children one year of age and under. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1985 Dec;53(4):517–23.

36. Girdhar A, Mishra B, Lavania RK, Bagga AK, Malaviya GN, Girdhar BK. Leprosy in infants--report of two cases. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1989 Jun;57(2):472–5.

37. Ghorpade A. Inoculation (tattoo) leprosy: a report of 31 cases. J Eur Acad Dermatol Venereol JEADV. 2002 Sep;16(5):494–9.

38. Bratschi MW, Steinmann P, Wickenden A, Gillis TP. Current knowledge on *Mycobacterium leprae* transmission: a systematic literature review. Lepr Rev. 2015 Jun;86(2):142–55.

39. Scollard DM. Pathogenesis and Pathology of Leprosy. In: The

International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

40. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The Continuing Challenges of Leprosy. Clin Microbiol Rev. 2006 Apr;19(2):338–81.

41. Vijayan J, Wilder-Smith EP. Neurological Manifestations of Leprosy. In: The International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

42. Shepard CC. The first decade in experimental leprosy. Bull World Health Organ. 1971;44(6):821–7.

43. Pattyn SR. The problem of cultivation of *Mycobacterium leprae.* Bull World Health Organ. 1973;49(4):403–10.

44. Shepard CC. The experimental disease that follows the injection of human leprosy bacilli into foot-pads of mice. J Exp Med. 1960 Sep 1;112(3):445–54.

45. Levy L, Ji B. The mouse foot-pad technique for cultivation of *Mycobacterium leprae*. Lepr Rev. 2006 Mar;77(1):5–24.

46. Storrs EE, Walsh GP, Burchfield HP, Binford CH. Leprosy in the Armadillo: New Model for Biomedical Research. Science. 1974 Mar 1;183(4127):851–2.

47. Balamayooran G, Pena M, Sharma R, Truman RW. The armadillo as an animal model and reservoir host for *Mycobacterium leprae*. Clin Dermatol. 2015 Jan;33(1):108–15.

48. Penna M T, Sharma R, Truman RW. The Armadillo Model for Leprosy. In: The International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

49. Modlin RL. The innate immune response in leprosy. Curr Opin Immunol. 2010 Feb;22(1):48–54.

50. Sousa JR, Sotto MN, Quaresma JAS. Leprosy as a complex infection: breakdown of the Th1 and Th2 immune paradigm in the immunopathogenesis of the disease. Front Immunol. 2017;8(1635).

51. Pinheiro RO, de Souza Salles J, Sarno EN, Sampaio EP. *Mycobacterium leprae*–host-cell interactions and genetic determinants in leprosy: an overview. Future Microbiol. 2011 Feb 1;6(2):217–30.

52. Geluk A. Challenges in immunodiagnostic tests for leprosy. Expert Opin Med Diagn. 2013 May;7(3):265–74.

53. Fitness J, Tosh K, Hill AVS. Genetics of susceptibility to leprosy. Genes Immun. 2002 Dec;3(8):441–53.

54. Fernando SL, Britton WJ. Genetic susceptibility to mycobacterial disease in humans. Immunol Cell Biol. 2006 Apr;84(2):125.

55. Sharma R, Lahiri R, Scollard DM, Pena M, Williams DL, Adams LB, et al. The armadillo: a model for the neuropathy of leprosy and potentially other neurodegenerative diseases. Dis Model Mech. 2013 Jan;6(1):19–24.

56. Abel L, Sánchez FO, Oberti J, Thuc NV, Hoa LV, Lap VD, et al. Susceptibility to leprosy is linked to the human NRAMP1 gene. J Infect Dis. 1998 Jan;177(1):133–45.

57. Mensah-Awere D, Bratschi MW, Steinmann P, Fairley JK, Gillis TP. Symposium Report: Developing Strategies to Block the Transmission of Leprosy. Lepr Rev. 2015 Jun;86(2):156–64.

58. Dunlap MD, Khader SA. Dancing with the Stars: Phenolic Glycolipids Partners with Macrophages. Cell Host Microbe. 2017 Sep 13;22(3):249–51.

59. Madigan CA, Cambier CJ, Kelly-Scumpia KM, Scumpia PO, Cheng T-Y, Zailaa J, et al. A Macrophage Response to *Mycobacterium leprae*

Phenolic Glycolipid Initiates Nerve Damage in Leprosy. Cell. 2017 Aug 24;170(5):973–85.e10.

60. Toledo-Pinto D, Gomes T, Ferreira ABR, Ribeiro-Alves M, Rodrigues LS, Batista-Silva LR, et al. STING-Dependent 2′-5′ Oligoadenylate Synthetase–Like Production Is Required for Intracellular *Mycobacterium leprae* Survival. J Infect Dis. 2016 Jul 15;214(2):311–20.

61. Cambier CJ, O'Leary SM, O'Sullivan MP, Keane J, Ramakrishnan L. Phenolic Glycolipid Facilitates Mycobacterial Escape from Microbicidal Tissue-Resident Macrophages. Immunity. 2017 Sep 19;47(3):552–65.e4.

62. Williams DL, Gillis TP, Fiallo P, Job CK, Gelber RH, Hill C, et al. Detection of *Mycobacterium leprae* and the potential for monitoring antileprosy drug therapy directly from skin biopsies by PCR. Mol Cell Probes. 1992 Oct 1;6(5):401–10.

63. Clark-Curtiss JE, Docherty MA. A Species-Specific Repetitive Sequence in *Mycobacterium leprae* DNA. J Infect Dis. 1989 Jan 1;159(1):7–15.

64. Honore N, Cole ST. Molecular basis of rifampin resistance in *Mycobacterium leprae*. Antimicrob Agents Chemother. 1993 Mar;37(3):414–8.

65. Fsihi H, Cole ST. The *Mycobacterium leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel polA locus associated with genomic variability. Mol Microbiol. 1995 Jun;16(5):909–19.

66. Eiglmeier K, Fsihi H, Heym B, Cole ST. On the catalase-peroxidase gene, katG, of *Mycobacterium leprae* and the implications for treatment of leprosy with isoniazid. FEMS Microbiol Lett. 1997 Apr 15;149(2):273–8.

67. Woods SA, Cole ST. A family of dispersed repeats in *Mycobacterium leprae*. Mol Microbiol. 1990 Oct;4(10):1745–51.

68. Grosskinsky CM, Jacobs WR, Clark-Curtiss JE, Bloom BR. Genetic relationships among *Mycobacterium leprae*, *Mycobacterium tuberculosis*, and candidate leprosy vaccine strains determined by DNA hybridization: identification of an *M. leprae*-specific repetitive sequence. Infect Immun. 1989 May 1;57(5):1535–41.

69. Williams DL, Gillis TP, Portaels F. Geographically distinct isolates of *Mycobacterium leprae* exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. Mol Microbiol. 1990 Oct;4(10):1653–9.

70. Donoghue HD, Holton J, Spigelman M. PCR primers that can detect low levels of *Mycobacterium leprae* DNA. J Med Microbiol. 2001;50(2):177–82.

71. Hartskeerl RA, De Wit MYL, Klatser PR. Polymerase Chain Reaction for the Detection of *Mycobacterium leprae*. Microbiology. 1989;135(9):2357–64.

72. Woods SA, Cole ST. A rapid method for the detection of potentially viable *Mycobacterium leprae* in human biopsies: a novel application of PCR. FEMS Microbiol Lett. 1989 Dec;53(3):305–9.

73. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001 Feb 22;409(6823):1007–11.

74. Singh P, Cole ST. *Mycobacterium leprae*: genes, pseudogenes and genetic diversity. Future Microbiol. 2011 Jan;6(1):57–71.

75. Gómez-Valero L, Rocha EPC, Latorre A, Silva FJ. Reconstructing the ancestor of *Mycobacterium leprae*: the dynamics of

gene loss and genome reduction. Genome Res. 2007 Aug;17(8):1178–85.

76. Cole ST, Supply P, Honoré N. Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. Lepr Rev. 2001 Dec;72(4):449–61.

77. Tufariello JM, Kerantzas CA, Vilchèze C, Calder RB, Nordberg EK, Fischer JA, et al. The Complete Genome Sequence of the Emerging Pathogen *Mycobacterium haemophilum* Explains Its Unique Culture Requirements. mBio. 2015 Dec 31;6(6):e01313–5.

78. Kimura M, Sakamuri RM, Groathouse NA, Rivoire BL, Gingrich D, Krueger-Koplin S, et al. Rapid variable-number tandem-repeat genotyping for *Mycobacterium leprae* clinical specimens. J Clin Microbiol. 2009 Jun 1;47(6):1757–66.

79. Young SK, Ponnighaus JM, Jain S, Lucas S, Suneetha S, Lockwood DNJ, et al. Use of Short Tandem Repeat Sequences to Study *Mycobacterium leprae* in Leprosy Patients in Malawi and India. PLoS Negl Trop Dis. 2008 Apr 9;2(4):e214.

80. Monot M, Honore N, Baliere C, Ji B, Sow S, Brennan PJ, et al. Are Variable-Number Tandem Repeats Appropriate for Genotyping *Mycobacterium leprae*? J Clin Microbiol. 2008 Jul;46(7):2291–7.

81. Da Silva Rocha A, Cunha dos Santos AA, Pignataro P, Nery JA, de Miranda AB, Soares DF, et al. Genotyping of *Mycobacterium leprae* from Brazilian leprosy patients suggests the occurrence of reinfection or of bacterial population shift during disease relapse. J Med Microbiol. 2011 Oct;60(Pt 10):1441–6.

82. Fontes ANB, Sakamuri RM, Baptista IMFD, Ura S, Moraes MO, Martínez AN, et al. Genetic diversity of *Mycobacterium leprae* isolates from Brazilian leprosy patients. Lepr Rev. 2009 Sep;80(3):302–15.

83. Hall BG, Salipante SJ. Molecular Epidemiology of *Mycobacterium leprae* as Determined by Structure-Neighbor Clustering. J Clin Microbiol. 2010 Jun;48(6):1997–2008.

84. Monot M, Honoré N, Garnier T, Araoz R, Coppée J-Y, Lacroix C, et al. On the origin of leprosy. Science. 2005 May 13;308(5724):1040–2.

85. Monot M, Honoré N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009 Dec;41(12):1282–9.

86. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013 Jul 12;341(6142):179–83.

87. Jacob JT, Franco-Paredes C. The Stigmatization of Leprosy in India and Its Impact on Future Approaches to Elimination and Control. PLoS Negl Trop Dis. 2008 Jan 30;2(1).

88. Secco RGCD, França K, Castillo D, AlHarbi M, Lotti T, Fioranelli M, et al. A synopsis of the history of Hansen's disease. Wien Med Wochenschr. 2017 Oct 1;167(1):27–30.

89. Amundson R, Ruddle-Miyamoto A. A Wholesome Horror: The Stigmas of Leprosy in 19th Century Hawaii. Disabil Stud Q. 2010 Aug 24;30(3/4).

90. Misch EA, Berrington WR, Vary JC, Hawn TR. Leprosy and the Human Genome. Microbiol Mol Biol Rev MMBR. 2010 Dec;74(4):589–620.

91. Fontilles (Spain) | International Leprosy Association History of Leprosy [Internet]. [cited 2017 Dec 11]. Available from: http://leprosyhistory.org/geographical_region/site/fontilles

92. Peters RMH, Zweekhorst MBM, Brakel WH van, Bunders JFG, Irwanto. "People like me don"t make things like that': Participatory video as a method for reducing leprosy-related stigma. Glob Public Health. 2016 Jul 2;11(5-6):666–82.

93. Dako-Gyeke M. Courtesy stigma: A concealed consternation among caregivers of people affected by leprosy. Soc Sci Med 1982. 2017 Nov 20;196:190–6.

94. Sarkar R, Pradhan S. Leprosy and women. Int J Womens Dermatol. 2016 Dec;2(4):117–21.

95. Price VG. Factors preventing early case detection for women affected by leprosy: a review of the literature. Glob Health Action. 2017 Dec;10(sup2):1360550.

96. Peters ES, Eshiet AL. Male-female (sex) differences in leprosy patients in south eastern Nigeria: females present late for diagnosis and treatment and have higher rates of deformity. Lepr Rev. 2002 Sep;73(3):262–7.

97. Noordeen SK. History of chemotherapy of leprosy. Clin Dermatol. 2016 Jan 1;34(1):32–6.

98. Williams DL, Spring L, Harris E, Roche P, Gillis TP. Dihydropteroate synthase of *Mycobacterium leprae* and dapsone resistance. Antimicrob Agents Chemother. 2000 Jun;44(6):1530–7.

99. Gelber RH. Progress in the chemotherapy of leprosy: Status, issues and prospects. In: Progress in Drug Research. Birkhäuser Basel; 1990. p. 421–45.

100. Shepard CC, Chang YT. Activity of antituberculosis drugs against *Mycobacterium leprae*. Int J Lepr. 1964 Sep;32:260–71.

101. Shepard CC. Chemotherapy of leprosy. Annu Rev Pharmacol. 1969;9:37–50.

102. Shepard CC, Levy L, Fasal P. Rapid bactericidal effect of rifampin on *Mycobacterium leprae*. Am J Trop Med Hyg. 1972 Jul;21(4):446–9.

103. WHO. Chemotherapy of leprosy. 1994. Report No.: 847.

104. Smith CS, Aerts A, Saunderson P, Kawuma J, Kita E, Virmond M. Multidrug therapy for leprosy: a game changer on the path to elimination. Lancet Infect Dis. 2017 Sep;17(9):e293–7.

105. Kar HK, Gupta R. Treatment of leprosy. Clin Dermatol. 2015 Jan 1;33(1):55–65.

106. Eichelmann K, González González SE, Salas-Alanis JC, Ocampo-Candiani J. Leprosy. An Update: Definition, Pathogenesis, Classification, Diagnosis, and Treatment. Actas Dermo-Sifiliográficas Engl Ed. 2013 Sep;104(7):554–63.

107. Virmond M, Joshua J, Solomon S, Duerksen F. Surgical aspects in leprosy. In: The International Textbook of Leprosy. Scollard DM, Gillis TP.

108. Pethe K, Bifani P, Jang J, Kang S, Park S, Ahn S, et al. Discovery of Q203, a potent clinical candidate for the treatment of tuberculosis. Nat Med. 2013 Sep;19(9):1157–60.

109. Kalia NP, Hasenoehrl EJ, Ab Rahman NB, Koh VH, Ang MLT, Sajorda DR, et al. Exploiting the synthetic lethality between terminal respiratory oxidases to kill *Mycobacterium tuberculosis* and clear host infection. Proc Natl Acad Sci U S A. 2017 Jul 11;114(28):7426–31.

110. Juréen P, Werngren J, Toro J-C, Hoffner S. Pyrazinamide Resistance and pncA Gene Mutations in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2008 May;52(5):1852–4.

111. Hoagland DT, Liu J, Lee RB, Lee RE. New agents for the treatment of drug-resistant *Mycobacterium tuberculosis*.

Adv Drug Deliv Rev. 2016 Jul 1;102:55–72.

112. Andries K, Villellas C, Coeck N, Thys K, Gevers T, Vranckx L, et al. Acquired Resistance of *Mycobacterium tuberculosis* to Bedaquiline. PLoS ONE. 2014 Jul 10;9(7).

113. Foo CS-Y, Lechartier B, Kolly GS, Boy-Röttger S, Neres J, Rybniker J, et al. Characterization of DprE1-Mediated Benzothiazinone Resistance in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2016 Oct 21;60(11):6451–9.

114. Xavier AS, Lakshmanan M. Delamanid: A new armor in combating drug-resistant tuberculosis. J Pharmacol Pharmacother. 2014;5(3):222–4.

115. Manjunatha UH, Lahiri R, Randhawa B, Dowd CS, Krahenbuhl JL, Barry CE. *Mycobacterium leprae* Is Naturally Resistant to PA-824. Antimicrob Agents Chemother. 2006 Oct;50(10):3350–4.

116. Saunderson PR. Drug-resistant M leprae. Clin Dermatol. 2016 Jan 1;34(1):79–81.

117. Gillis TP, Williams DL. Dapsone resistance in *Mycobacterium leprae*. Lepr Rev. 2000 Dec;71 Suppl:S91–5.

118. Cambau E, Perani E, Guillemin I, Jamet P, Ji B. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. Lancet. 1997 Jan 11;349(9045):103–4.

119. Hatfull GF, Jacobs, Jr WR. Molecular Genetics of Mycobacteria. 2nd ed. American Society for Microbiology Press; 2014. 1474 p.

120. Swanson RV, Adamson J, Moodley C, Ngcobo B, Ammerman NC, Dorasamy A, et al. Pharmacokinetics and pharmacodynamics of clofazimine in a mouse model of tuberculosis. Antimicrob Agents Chemother. 2015;59(6):3042–51.

121. Cholo MC, Steel HC, Fourie PB, Germishuizen WA, Anderson R. Clofazimine: current status and future prospects. J Antimicrob Chemother. 2011 Oct 20;dkr444.

122. Lechartier B, Cole ST. Mode of Action of Clofazimine and Combination Therapy with Benzothiazinones against *Mycobacterium tuberculosis*. Antimicrob Agents Chemother. 2015 Aug;59(8):4457–63.

123. Maeda S, Matsuoka M, Nakata N, Kai M, Maeda Y, Hashimoto K, et al. Multidrug Resistant *Mycobacterium leprae* from Patients with Leprosy. Antimicrob Agents Chemother. 2001 Dec;45(12):3635–9.

124. Cambau E, Bonnafous P, Perani E, Sougakoff W, Ji B, Jarlier V. Molecular Detection of Rifampin and Ofloxacin Resistance for Patients Who Experience Relapse of Multibacillary Leprosy. Clin Infect Dis. 2002;34(1):39–45.

125. Matsuoka M. Drug resistance in leprosy. Jpn J Infect Dis. 2010 Jan;63(1):1–7.

126. Yokoyama K, Kim H, Mukai T, Matsuoka M, Nakajima C, Suzuki Y. Impact of amino acid substitutions in B subunit of DNA gyrase in *Mycobacterium leprae* on fluoroquinolone resistance. PLoS Negl Trop Dis. 2012;6(10):e1838.

127. Cambau E, Chauffour-Nevejans A, Tejmar-Kolar L, Matsuoka M, Jarlier V. Detection of Antibiotic Resistance in Leprosy Using GenoType LepraeDR, a Novel Ready-To-Use Molecular Test. PLoS Negl Trop Dis [Internet]. 2012 Jul 31 [cited 2013 Nov 4];6(7). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3409109/

128. Matsuoka M. Global surveillance system to monitor the development of drug resistance in *Mycobacterium leprae*. Res Rep Trop Med. 2015 Nov;75.

129. WHO. Antimicrobial resistance in leprosy: report of a global consultation. Kathmandu; 2016.

130. Da Silva Rocha A, Cunha M das G, Diniz LM, Salgado C, Aires MAP, Nery JA, et al. Drug and multidrug resistance among *Mycobacterium leprae* isolates from Brazilian relapsed leprosy patients. J Clin Microbiol. 2012 Jun;50(6):1912–7.

131. Lavania M, Singh I, Turankar RP, Ahuja M, Pathak V, Sengupta U, et al. Molecular detection of multi drug resistant *Mycobacterium leprae* from Indian leprosy patients. J Glob Antimicrob Resist. 2017 Oct 31;

132. Guerrero MI, Colorado CL, Torres JF, León CI. Is drug-resistant *Mycobacterium leprae* a real cause for concern?: First approach to molecular monitoring of multibacillary Colombian patients with and without previous leprosy treatment. Bioméd Rev Inst Nac Salud. 2014 Apr;34 Suppl 1:137–47.

133. Williams DL, Hagino T, Sharma R, Scollard D. Primary Multidrug-Resistant Leprosy, United States. Emerg Infect Dis. 2013 Jan;19(1):179–81.

134. Williams DL, Lewis C, Sandoval FG, Robbins N, Keas S, Gillis TP, et al. Drug Resistance in Patients With Leprosy in the United States. Clin Infect Dis. 2014 Jan 1;58(1):72–3.

135. Liu D, Zhang Q, Sun Y, Wang C, Zhang Y, Fu X, et al. Drug resistance in *Mycobacterium leprae* from patients with leprosy in China. Clin Exp Dermatol. 2015 May 1;n/a – n/a.

136. Reed SG, Duthie MS. Vaccine for prevention of leprosy. In: The International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

137. Convit J, Sampson C, Zúñiga M, Plata J, Silva J, Molina J, et al. Immunoprophylactic trial with combined *Mycobacterium leprae*/BCG vaccine against leprosy: preliminary results. The Lancet. 1992 Feb 22;339(8791):446–50.

138. Kumar A, Parkash O, Girdhar BK. Analysis of Antigens of *Mycobacterium leprae* by Interaction to Sera IgG, IgM, and IgA Response to Improve Diagnosis of Leprosy. BioMed Res Int. 2014;2014.

139. Duthie MS, Sampaio LH, Oliveira RM, Raman VS, O'Donnell J, Bailor HR, et al. Development and pre-clinical assessment of a 73kD chimeric fusion protein as a defined sub-unit vaccine for leprosy. Vaccine. 2013 Jan 21;31(5):813–9.

140. Kamal R, Pathak V, Kumari A, Natrajan M, Katoch K, Kar HK. Addition of Mycobacterium indicus pranii vaccine as an immunotherapeutic to standard chemotherapy in borderline leprosy: a double-blind study to assess clinical improvement (preliminary report). Br J Dermatol. 2017;176(5):1388–9.

141. Lavania M, Jadhav R, Turankar RP, Singh I, Nigam A, Sengupta U. Genotyping of *Mycobacterium leprae* strains from a region of high endemic leprosy prevalence in India. Infect Genet Evol. 2015 Dec 1;36(Supplement C):256–61.

142. Donoghue HD, Michael Taylor G, Marcsik A, Molnár E, Pálfi G, Pap I, et al. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. Infect Genet Evol [Internet]. [cited 2015 Feb 16]; Available from: http://www.sciencedirect.com/science/article/pii/S1567134815000441

143. Bos KI, Schuenemann VJ, Golding GB, Burbano HA, Waglechner N, Coombes BK, et al. A draft genome of Yersinia pestis from victims of the Black Death. Nature. 2011 Oct 27;478(7370):506–10.

144. Falush D, Wirth T, Linz B, Pritchard JK, Stephens M, Kidd M, et al. Traces of Human Migrations in Helicobacter pylori Populations. Science. 2003 Mar 7;299(5612):1582–5.

145. Whatmore AM. Ancient-Pathogen Genomics: Coming of Age? mBio. 2014 Oct 31;5(5):e01676–14.

146. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, et al. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014 Oct 23;514(7523):494–7.

147. Zink AR, Reischl U, Wolf H, Nerlich AG. Molecular analysis of ancient microbial infections. FEMS Microbiol Lett. 2002 Aug 1;213(2):141–7.

148. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC Genomics. 2014 Apr 8;15(1):270.

149. Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, et al. Molecular, ethno-spatial epidemiology of leprosy in China: Novel insights for tracing leprosy in endemic and non endemic provinces. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2013 Mar;14:361–8.

150. Han XY, Seo Y-H, Sizer KC, Schoberle T, May GS, Spencer JS, et al. A new Mycobacterium species causing diffuse lepromatous leprosy. Am J Clin Pathol. 2008 Dec;130(6):856–64.

151. Han XY, Sizer KC, Thompson EJ, Kabanja J, Li J, Hu P, et al. Comparative Sequence Analysis of *Mycobacterium leprae* and the New Leprosy-Causing *Mycobacterium lepromatosis*. J Bacteriol. 2009 Oct 1;191(19):6067–74.

152. Vera-Cabrera L, Escalante-Fuentes WG, Gomez-Flores M, Ocampo-Candiani J, Busso P, Singh P, et al. Case of diffuse lepromatous leprosy associated with "*Mycobacterium lepromatosis*." J Clin Microbiol. 2011 Dec;49(12):4366–8.

153. Gillis TP, Scollard DM, Lockwood DNJ. What is the evidence that the putative *Mycobacterium lepromatosis* species causes diffuse lepromatous leprosy? Lepr Rev. 2011 Sep;82(3):205–9.

154. Han XY, Sizer KC, Tan H-H. Identification of the leprosy agent *Mycobacterium lepromatosis* in Singapore. J Drugs Dermatol JDD. 2012 Feb;11(2):168–72.

155. Jessamine PG, Desjardins M, Gillis T, Scollard D, Jamieson F, Broukhanski G, et al. Leprosy-like illness in a patient with *Mycobacterium lepromatosis* from Ontario, Canada. J Drugs Dermatol JDD. 2012 Feb;11(2):229–33.

156. Han XY, Sizer KC, Velarde-Félix JS, Frias-Castro LO, Vargas-Ocampo F. The leprosy agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico. Int J Dermatol. 2012;51(8):952–9.

157. Carpenter CM. Murine Leprosy—Its Usefulness as an Experimental Infection. Ann N Y Acad Sci. 1951 Mar 1;54(1):101–6.

158. Courtin F, Huerre M, Fyfe J, Dumas P, Boschiroli ML. A case of feline leprosy caused by *Mycobacterium lepra*murium originating from the island of Kythira (Greece): diagnosis and treatment. J Feline Med Surg. 2007 Jun;9(3):238–41.

159. Rojas-Espinosa O. Murine Leprosy Revisited. 2009. 97 p.

160. Goulart IMB, Bernardes Souza DO, Marques CR, Pimenta VL, Gonçalves MA, Goulart LR. Risk and Protective Factors for Leprosy Development Determined by Epidemiological Surveillance of Household Contacts.

Clin Vaccine Immunol CVI. 2008 Jan;15(1):101–5.

161. Kumar B, Uprety S, Dogra. Clinical diagnosis of leprosy. In: The International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

162. Gelber RH, Balagon VF, Cellona RV. The relapse rate in MB leprosy patients treated with 2-years of WHO-MDT is not low. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 2004 Dec;72(4):493–500.

163. Girdhar BK, Girdhar A, Kumar A. Relapses in multibacillary leprosy patients: effect of length of therapy. Lepr Rev. 2000 Jun;71(2):144–53.

164. Oskam L, Dockrell HM, Brennan PJ, Gillis T, Vissa V, Richardus JH, et al. Molecular methods for distinguishing between relapse and reinfection in leprosy. Trop Med Int Health. 2008 Oct 1;13(10):1325–6.

165. Young D. Prospects for molecular epidemiology of leprosy. Lepr Rev. 2003 Mar;74(1):11–7.

166. Job CK. Leprosy--the source of infection and its mode of transmission. Lepr Rev. 1981 Dec;52 Suppl 1:69–76.

167. Kirchheimer AWF. Leprosy in the Nine-Banded Armadillo. Public Health Rep. 1975;90(6):483–5.

168. Walsh GP, Meyers WM, Binford CH. Naturally Acquired Leprosy in the NineBanded Armadillo: A Decade of Experience 1975 1985. J Leukoc Biol. 1986;40(5):645–56.

169. Donham KJ, Leininger JR. Spontaneous leprosy-like disease in a chimpanzee. J Infect Dis. 1977 Jul;136(1):132–6.

170. Gormus BJ, Wolf RH, Baskin GB, Ohkawa S, Gerone PJ, Walsh GP, et al. A second sooty mangabey monkey with naturally acquired leprosy: first reported possible monkey-to-monkey transmission. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1988 Mar;56(1):61–5.

171. Leininger JR, Donham KJ, Meyers WM. Leprosy in a chimpanzee. Postmortem lesions. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1980 Dec;48(4):414–21.

172. Sharma R, Singh P, Loughry WJ, Lockhart JM, Inman WB, Duthie MS, et al. Zoonotic leprosy in the southeastern United States. Emerg Infect Dis. 2015;21(12):2127–34.

173. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable Zoonotic Leprosy in the Southern United States. N Engl J Med. 2011;364(17):1626–33.

174. Taulman JF, Robbins LW. Range expansion and distributional limits of the nine-banded armadillo in the United States: an update of Taulman & Robbins (1996). J Biogeogr. 2014 Aug 1;41(8):1626–30.

175. Smith LL. The amazing armadillo : geography of a folk critter /. 1st ed. Austin : University of Texas Press,; 1984.

176. Truman R. Leprosy in wild armadillos. Lepr Rev. 2005 Sep;76(3):198–208.

177. Meredith A, Pozo JD, Smith S, Milne E, Stevenson K, McLuckie J. Leprosy in red squirrels in Scotland. Vet Rec. 2014 Sep 20;175(11):285–6.

178. Simpson V, Hargreaves J, Butler H, Blackett T, Stevenson K, McLuckie J. Leprosy in red squirrels on the Isle of Wight and Brownsea Island. Vet Rec. 2015 Aug 22;177(8):206–7.

179. Steinmann P, Reed SG, Mirza F, Hollingsworth TD, Richardus JH. Innovative tools and approaches to end the transmission of *Mycobacterium leprae*. Lancet Infect Dis. 2017 Sep 1;17(9):e298–305.

180. Al PEB et. Leprosy in Nonimmigrant Canadian Man without Travel outside North America, 2014 Volume 24, Number 1—January 2018 Emerging Infectious Disease journal CDC. Emerg Infect Dis. 2018;24:165–6.

181. Lavania M, Katoch K, Katoch VM, Gupta AK, Chauhan DS, Sharma R, et al. Detection of viable *Mycobacterium leprae* in soil samples: insights into possible sources of transmission of leprosy. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2008 Sep;8(5):627–31.

182. Adams LB, Lahiri R. Cultivation and Viability Determination of *Mycobacterium leprae*. In: The International Textbook of Leprosy. Scollard DM, Gillis TP. 2016.

183. Greub G, Raoult D. Microorganisms Resistant to Free-Living Amoebae. Clin Microbiol Rev. 2004 Apr 1;17(2):413–33.

184. Lahiri R, Krahenbuhl JL. The role of free-living pathogenic amoeba in the transmission of leprosy: a proof of principle. Lepr Rev. 2008 Dec;79(4):401–9.

185. Wheat WH, Casali AL, Thomas V, Spencer JS, Lahiri R, Williams DL, et al. Long-term Survival and Virulence of *Mycobacterium leprae* in Amoebal Cysts. PLoS Negl Trop Dis. 2014 Dec 18;8(12):e3405.

186. Manglani PR. Prevention of disability in leprosy. J Indian Med Assoc. 2004 Dec;102(12):680–3.

187. Sermrittirong S, Van Brakel WH, Bunbers-Aelen JFG. How to reduce stigma in leprosy--a systematic literature review. Lepr Rev. 2014 Sep;85(3):149–57.

188. Ridley DS, Hilson GR. A logarithmic index of bacilli in biopsies. I. Method. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 1967 Jun;35(2):184–6.

189. Job CK, Jayakumar J, Kearney M, Gillis TP. Transmission of Leprosy: A Study of Skin and Nasal Secretions of Household Contacts of Leprosy Patients Using PCR. Am J Trop Med Hyg. 2008 Mar 1;78(3):518–21.

190. Reis EM, Araujo S, Lobato J, Neves AF, Costa AV, Gonçalves MA, et al. *Mycobacterium leprae* DNA in peripheral blood may indicate a bacilli migration route and high-risk for leprosy onset. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis. 2014 May;20(5):447–52.

191. Lastória JC, de Abreu MAMM. Leprosy: a review of laboratory and therapeutic aspects Part 2. An Bras Dermatol. 2014;89(3):389–401.

192. Greenwood N, Fox AH. A comparison of methods for staining tubercle bacilli in histological sections. J Clin Pathol. 1973 Apr;26(4):253–7.

193. Bhardwaj K, Ghate S, Dhurat R. Detection of *Mycobacterium leprae* in tissue sections using auramine O fluorescent stain versus modified fite-faraco: A comparative study. Int J Infect Dis. 2016 Apr 1;45:388.

194. Elston DM, Liranzo MO, Scollard DM. Comparing the sensitivity of auramine-rhodamine fluorescence to polymerase chain reaction in the detection of *Mycobacterium leprae* in Fite-negative tissue sections. J Am Acad Dermatol. 2017 May 1;76(5):992–3.

195. Truman RW, Andrews PK, Robbins NY, Adams LB, Krahenbuhl JL, Gillis TP. Enumeration of *Mycobacterium leprae* Using Real-Time PCR. PLOS Negl Trop Dis. 2008 Nov 4;2(11):e328.

196. Azevedo M de CS, Ramuno NM, Fachin LRV, Tassa M, Rosa PS, Belone A de FF, et al. qPCR detection of *Mycobacterium leprae* in biopsies and slit skin smear of different leprosy clinical forms. Braz J Infect Dis.

197. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-Based Assays for Leprosy Diagnosis Directly in Clinical Specimens. PLoS Negl Trop Dis. 2011 Oct 11;5(10):e1354.

198. Hartskeerl RA, De Wit MYL, Klatser PR. Polymerase Chain Reaction for the Detection of *Mycobacterium leprae*. Microbiology. 1989;135(9):2357–64.

199. Turankar RP, Pandey S, Lavania M, Singh I, Nigam A, Darlong J, et al. Comparative evaluation of PCR amplification of RLEP, 16S rRNA, rpoT and Sod A gene targets for detection of *M. leprae* DNA from clinical and environmental samples. Int J Mycobacteriology. 2015 Mar;4(1):54–9.

200. Martinez AN, Talhari C, Moraes MO, Talhari S. PCR-Based Techniques for Leprosy Diagnosis: From the Laboratory to the Clinic. PLoS Negl Trop Dis. 2014 Apr 10;8(4).

201. Tyring SK, Lupi O, Hengge UR, editors. 22 Mycobacteria. In: Tropical Dermatology (Second Edition). Elsevier; 2017. p. 252–79.

202. Bernardes Filho F, Paula NA de, Leite MN, Abi-Rached TLC, Vernal S, Silva MB da, et al. Evidence of hidden leprosy in a supposedly low endemic area of Brazil. Mem Inst Oswaldo Cruz. 2017 Dec;112(12):822–8.

203. Bührer-Sékula S. PGL-I leprosy serology. Rev Soc Bras Med Trop. 2008 Jan;41:3–5.

204. Serrano-Coll H, Muñoz M, Camilo Beltrán J, Duthie MS, Cardona-Castro N. Anti-natural octyl disaccharide-leprosy IDRI diagnostic (NDO-LID) antibodies as indicators of leprosy reactions and neuritis. Trans R Soc Trop Med Hyg. 2017 Mar 1;111(3):125–31.

205. Geluk A, van Meijgaarden KE, Franken KLMC, Subronto YW, Wieles B, Arend SM, et al. Identification and Characterization of the ESAT-6 Homologue of *Mycobacterium leprae* and T-Cell Cross-Reactivity with *Mycobacterium tuberculosis*. Infect Immun. 2002 May;70(5):2544–8.

206. Gey van Pittius NC, Warren RM, van Helden PD. ESAT-6 and CFP-10: What Is the Diagnosis? Infect Immun. 2002 Nov;70(11):6509–11.

207. Geluk A, van Meijgaarden KE, Franken KLMC, Wieles B, Arend SM, Faber WR, et al. Immunological crossreactivity of the *Mycobacterium leprae* CFP-10 with its homologue in *Mycobacterium tuberculosis*. Scand J Immunol. 2004 Jan;59(1):66–70.

208. Rendini T, Levis W. Quantiferon-Gold Tuberculosis Test Cannot Detect Latent Tuberculosis in Patients With Leprosy. Clin Infect Dis. 2015 Nov 1;61(9):1439–40.

209. Savage JA, Levis WR. The utility of QuantiFERON-TB Gold assay in patients with leprosy: A possible measure of anergy. J Am Acad Dermatol. 2014 Nov 1;71(5):996–7.

210. Van Hooij A, Tjon Kon Fat EM, Richardus R, van den Eeden SJF, Wilson L, de Dood CJ, et al. Quantitative lateral flow strip assays as User-Friendly Tools To Detect Biomarker Profiles For Leprosy. Sci Rep. 2016 Sep 29;6:34260.

211. Mirsaeidi M, Banoei MM, Winston BW, Schraufnagel DE. Metabolomics: Applications and Promise in Mycobacterial Disease. Ann Am Thorac Soc. 2015 Jul 21;12(9):1278–87.

212. Al-Mubarak R, Heiden JV, Broeckling CD, Balagon M, Brennan PJ, Vissa VD. Serum Metabolomics Reveals Higher Levels of Polyunsaturated Fatty Acids in Lepromatous Leprosy: Potential Markers for Susceptibility and Pathogenesis. PLoS Negl Trop Dis. 2011 Sep 6;5(9):e1303.

213. Mayboroda OA, van Hooij A, Derks R, van den Eeden SJF, Dijkman K, Khadge S, et al. Exploratory urinary metabolomics of type 1 leprosy reactions. Int J Infect Dis IJID Off Publ Int Soc Infect Dis. 2016 Apr;45:46–52.

214. Araujo D, S L, Vaas LAI, Ribeiro-Alves M, Geffers R, Mello FCQ, et al. Transcriptomic Biomarkers for Tuberculosis: Evaluation of DOCK9. EPHA4, and NPC2 mRNA Expression in Peripheral Blood. Front Microbiol. 2016;7.

215. Darboe F, Mbandi SK, Thompson EG, Fisher M, Rodo M, van Rooyen M, et al. Diagnostic performance of an optimized transcriptomic signature of risk of tuberculosis in cryopreserved peripheral blood mononuclear cells. Tuberculosis. 2018 Jan 1;108(Supplement C):124–6.

216. Berry MPR, Graham CM, McNab FW, Xu Z, Bloch SAA, Oni T, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature. 2010 Aug 19;466(7309):973–7.

217. Dupnik KM, Bair TB, Maia AO, Amorim FM, Costa MR, Keesen TSL, et al. Transcriptional changes that characterize the immune reactions of leprosy. J Infect Dis. 2015 May 15;211(10):1658–76.

218. Romero-Montoya M, Beltran-Alzate JC, Cardona-Castro N. Evaluation and Monitoring of *Mycobacterium leprae* Transmission in Household Contacts of Patients with Hansen's Disease in Colombia. PLoS Negl Trop Dis. 2017 Jan 23;11(1):e0005325.

219. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A prospective blood RNA signature for tuberculosis disease risk. Lancet Lond Engl. 2016 Jun 4;387(10035):2312–22.

220. Zak DE, Penn-Nicholson A, Scriba TJ, Thompson E, Suliman S, Amon LM, et al. A blood RNA signature for tuberculosis disease risk: a prospective cohort study. The Lancet. 2016 Jun 4;387(10035):2312–22.

221. Fonseca JD, Knight GM, McHugh TD. The complex evolution of antibiotic resistance in *Mycobacterium tuberculosis*. Int J Infect Dis. 2015 Mar;32:94–100.

222. Günther G. Multidrug-resistant and extensively drug-resistant tuberculosis: a review of current concepts and future challenges. Clin Med Lond Engl. 2014 Jun;14(3):279–85.

223. Thompson EG, Du Y, Malherbe ST, Shankar S, Braun J, Valvo J, et al. Host blood RNA signatures predict the outcome of tuberculosis treatment. Tuberculosis. 2017 Dec 1;107(Supplement C):48–58.

Chapter 2. : Toward an *M. leprae* complex

Chapter 2.1 – Insight into the evolution and origin of bacilli from the genome sequence of *Mycobacterium lepromatosis*

Pushpendra Singh[a,*], Andrej Benjak[a,*], Verena J. Schuenemann[b], Alexander Herbig[b], Charlotte Avanzi[a], Philippe Busso[a], Kay Nieselt[c], Johannes Krause[b,d,e], Lucio Vera-Cabrera[f], and Stewart T. Cole[a,#]

[a]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Switzerland;
[b]Institute for Archaeological Sciences, University of Tübingen, Germany
[c]Center for Bioinformatics Tübingen, University of Tübingen, Germany
[d]Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, Germany
[e]Max Planck Institute for the Science of Human History, Jena, Germany
[f]Laboratorio Interdisciplinario de Investigación Dermatológica, Servicio de Dermatología, Hospital Universitario,Universidad Autónoma de Nuevo León, 64460 Monterrey, Nuevo León, Mexico

[*] These authors contributed equally to this work
[#] Corresponding author

2015. PNAS 112(14):4459-64

Contributions: second genome sequencing, PCR screening and manuscript preparation

This is a pre-copyedited, author-produced version of an article accepted for publication in PNAS following peer review. The version of record is available online at:
doi: 10.1073/pnas.1421504112

Abstract

Mycobacterium lepromatosis is an uncultured human pathogen associated with diffuse lepromatous leprosy and a reactional state known as Lucio's phenomenon. By using deep sequencing with and without DNA enrichment, we obtained the near-complete genome sequence of *M. lepromatosis* present in a skin biopsy from a Mexican patient, and compared it with that of *Mycobacterium leprae*, which has undergone extensive reductive evolution. The genomes display extensive synteny and are similar in size (~3.27 Mb). Protein-coding genes share 93% nucleotide sequence identity, whereas pseudogenes are only 82% identical. The events that led to pseudogenization of 50% of the genome likely occurred before divergence from their most recent common ancestor (MRCA), and both *M. lepromatosis* and *M. leprae* have since accumulated new pseudogenes or acquired specific deletions. Functional comparisons suggest that *M. lepromatosis* has lost several enzymes required for amino acid synthesis whereas *M. leprae* has a defective heme pathway. *M. lepromatosis* has retained all functions required to infect the Schwann cells of the peripheral nervous system and therefore may also be neuropathogenic. A phylogeographic survey of 227 leprosy biopsies by differential PCR revealed that 221 contained *M. leprae* whereas only six, all from Mexico, harbored *M. lepromatosis*. Phylogenetic comparisons indicate that *M. lepromatosis* is closer than *M. leprae* to the MRCA, and a Bayesian dating analysis suggests that they diverged from their MRCA approximately 13.9 Mya. Thus, despite their ancient separation, the two leprosy bacilli are remarkably conserved and still cause similar pathologic conditions.

Significance

Leprosy was thought to be exclusively caused by infection of humans by *Mycobacterium leprae*. In 2008, Han et al. proposed that *Mycobacterium lepromatosis*, a separate unculturable species, might be responsible for a rare yet severe form of the disease called diffuse lepromatous leprosy. Here, by using comparative genomics, we show that the two species are very closely related and derived from a common ancestor that underwent genome downsizing and gene decay. Since their separation 13.9 Mya, the two species have continued to lose genes, but from different regions of the genome, and *M. leprae* appears to be more recent. In a recent phylogeographic survey, by using differential PCR, we found that *M. lepromatosis* was scarce and restricted to patients from Mexico.

Introduction

Nearly a quarter million new case of leprosy (Hansen's disease) are still recorded annually worldwide despite a remarkable decrease in prevalence in the past decade (1). Leprosy primarily affects the skin, peripheral nerves, and eyes, and manifests as a spectrum of diverse clinical forms varying in bacillary load and often accompanied by painful immunological reactions (2–4). A severe form of leprosy known as diffuse lepromatous leprosy (DLL) that is common in western Mexico and the Caribbean region, first described by Lucio and Alvarado in 1852 (5), is referred to as Lucio's leprosy. Such cases account for a sizable proportion (more than 20%) of all leprosy cases in western Mexico (5–8), Cuba (9), and Costa Rica (10) but are rarely reported elsewhere. In 1948, Latapi and Zamora noted that DLL cases had no dermal nodules and were characterized by a generalized and diffuse infiltration of the skin by histiocytes and acid-fast bacilli causing an appearance of swollen or "spotted" skin, which they termed "pure and primitive diffuse lepromatosis." In addition, they reported that some patients developed acute necrotic skin reactions, "erythema necroticans," and differentiated this condition as Lucio's phenomenon (6).

The most notable clinical feature of DLL and Lucio's phenomenon is the diffuse mycobacterial invasion of endothelial cells surrounding small vessels, often leading to vascular occlusion (8). The initial cyanotic lesions, caused by poor blood supply and ischemia, gradually evolve into black necrotic lesions (11, 12). Hence, these cases are often associated with long-term morbidity (8) as well as a higher number of fatalities if not managed adequately (5, 13). Lucio's phenomenon is usually observed among untreated or inadequately treated nonnodular DLL cases 1–3 y after the

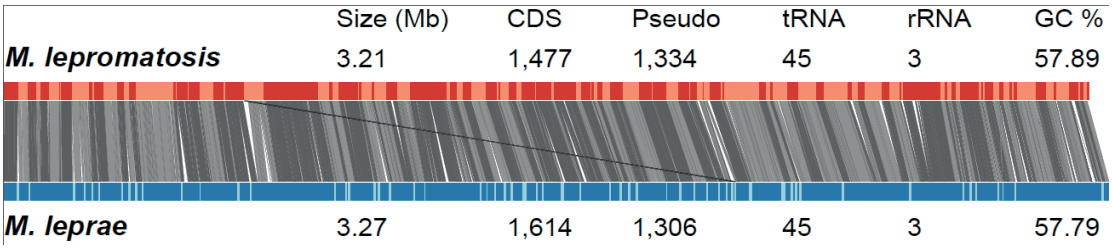


Figure 1: Genome synteny and salient features of *M. leprae* and *M. lepromatosis*. The 126 contigs of *M. lepromatosis* are distinguished with red and orange colors. Links between M. leprae and M. lepromatosis are BLAST hits, with two shades of gray to distinguish individual contigs. White stripes indicate no BLAST hits and account for 5–6% of each genome's specific sequences. Light blue stripes indicate dispersed repeats in M. leprae. Black line indicates a confirmed structural variation between *M. leprae* and *M. lepromatosis*

appearance of their symptoms. Although rare, Lucio's phenomenon has also been reported among other forms of lepromatous leprosy (6). Multidrug therapy (MDT) is currently the treatment recommended by the World Health Organization for all forms of leprosy, including DLL and Lucio's phenomenon (14, 15). Until recently, *Mycobacterium leprae* was considered the sole causative agent of all forms of leprosy, including Lucio's phenomenon, which is often referred to as a "form of leprosy reaction" (13, 16). The genome of *M. leprae* has undergone reductive evolution, with approximately half occupied by pseudogenes (17), and also displays remarkably low levels of genetic diversity (18–20).

In 2008, a new mycobacterial species named *Mycobacterium lepromatosis* was identified in a liver autopsy specimen from a homeless Mexican who died with DLL in Arizona (21). Since then, this species has been identified by PCR-based sequencing in several Mexican patients (22) as well as in individual cases from Singapore (23) and Canada (24). In addition, clinical presentations resembling Lucio's phenomenon have been reported elsewhere, i.e., Brazil (25, 26), India (27–29), Iran (30), and Malaysia (31); however, molecular confirmation of the mycobacterial agent was not carried out. In addition, several cases of mixed infection have been reported whereby both *M. lepromatosis* and *M. leprae* were detected (22), which undermines confidence in *M. lepromatosis* being the causative agent of DLL. Knowledge about the biology and pathogenesis of *M. lepromatosis* is limited because this species remains uncultivated (21).

To date, the DNA sequences of 22.8 kb of selected PCR fragments from *M. lepromatosis* are known, and these were sufficient to reveal striking sequence similarity to *M. leprae* and a close phylogenetic relationship, but this preliminary analysis provided little biological insight. Thus, at this stage, genome sequencing is the most efficient approach to investigate *M. lepromatosis*. After the first description of *M. lepromatosis* by Han et al. in 2008 (21), we reported independent confirmation of this species in a biopsy specimen from a DLL case (Mx1-22A) from Monterrey, Mexico (32). Here, we combined various DNA enrichment approaches and deep sequencing to unveil the genome of *M. lepromatosis* directly from the archived biopsy specimen from this patient. Genome-wide comparison of *M. lepromatosis* and *M. leprae* provides deeper insight into the biology of *M. lepromatosis* and discloses the evolutionary history of these two closely related but clearly distinct species.

Results and Discussion

Genome Sequencing, Assembly, and Analysis.

Because *M. lepromatosis* cannot be cultured *in vitro* and an animal model is not yet available, the only source of its DNA is infected human tissue. Biopsy specimen Mx1-22A was used for DNA extraction and Illumina library preparation as described previously (20). To overcome the problem of host DNA, we used two methods to enrich *M. lepromatosis* DNA: whole-genome array capture using the *M. leprae* genome as bait (20) and removal of human DNA by

hybridization with a human genomic DNA bait library (*Materials and Methods*). Illumina sequencing of the enriched as well as the original libraries provided 55-fold coverage, which was more than sufficient for the de novo whole-genome assembly (Table S1). The genome assembly was obtained by relaxing the assembler program's parameters (*Materials and Methods*) to account for the extremely biased read coverage from the array capture library. Contigs were considerably longer compared with those obtained with default assembly parameters, but at the cost of a higher chance of misassemblies. To avoid assembly errors, we split those contigs showing disrupted synteny with *M. leprae* so that, from the initial 110 contigs of the de novo assembly, we obtained a final set of 126 contigs. Most inconsistencies in the contigs were observed around areas of repetitive DNA, with some exceptions. One such exception was confirmed by PCR sequencing (as detailed later), proving that there is at least one instance of genome rearrangement between *M. leprae* and *M. lepromatosis*. However, we consider that the overall level of rearrangement is low, given that only a few contigs showed breaks in synteny with the reference genome and the GC skew of the "syntenic" version of the *M. lepromatosis* genome is virtually identical to that of *M. leprae* (Fig. S1). A total of 3,206,741 bases of the *M. lepromatosis* Mx1-22A genome were represented in the 126 contigs, and, with one exception, these all aligned to the 3.27-Mb circular genome sequence of the Tamil Nadu (TN) reference strain of *M. leprae* (17). The exception was a 2.3-kb contig bearing five mycobacterial pseudogenes. A graphical comparison revealing genome-wide synteny and key features of the two genomes is presented in Fig. 1. In brief, the genome of *M. lepromatosis* appears to harbor at least 1,477 genes encoding proteins [i.e., coding DNA sequence (CDS)] and 1,334 pseudogenes.

Repetitive DNA.

Dispersed repeats were the major cause of breaks in our *de novo* assembly. Interestingly, in most cases, the locations of these repeats correspond to those of the four main families of repetitive DNA in *M. leprae*: RLEP (37 copies), REPLEP (15 copies), LEPREP (8 copies), and LEPRPT (5 copies) (33). These shared 75–90% sequence identity, in segments as much as 350 nt in length, with the most conserved corresponding to LEPREP and LEPRPT repeats and the lowest homology found for the REPLEP repeats. From this analysis, it is evident that the four repeat families were present in the most recent common ancestor (MRCA) of *M. leprae* and *M. lepromatosis* and that the levels of conservation between the repeats in the two species are proportional to their copy number. No additional repetitive DNA was detected in *M. lepromatosis*.

A second draft genome.

Using the *M. lepromatosis* genome sequence, we designed specific PCR primers and used them to screen ~260 biopsies from our collection, thereby identifying a second case of *M. lepromatosis* in another Mexican patient (Mx177) who presented with Lucio's phenomenon. This sample, which contained no *M. leprae* DNA, was shotgun-sequenced, without any enrichment, and the resultant sequences were mapped against the Mx1-22A genome assembly. Despite shallow coverage of the Mx177 sample (80% of the genome at an average coverage of 5× after excluding duplicate reads), we found only 12 SNPs in a 2-Mb alignment where SNP calling was feasible. This very low SNP frequency (1 in 167 kb) is reminiscent of the similarly low genetic diversity of *M. leprae* strains from the same geographical area (34) and in a set of worldwide *M. leprae* genomes in general (20).

Synteny and conservation of the M. lepromatosis and M. leprae Genomes.

Ninety-four percent of the *M. lepromatosis* genome assembly could be aligned to the *M. leprae* TN genome, and there appears to be near-perfect collinearity and synteny, with 92% of the genes and pseudogenes shared (Fig. 1). Details of (pseudo)genes that have been deleted are provided in Dataset S1. In *M. leprae*, there are tandemly arranged asparagine permease genes (*ML1304c ML1305c*; *ansP1 ansP2*), but only one of

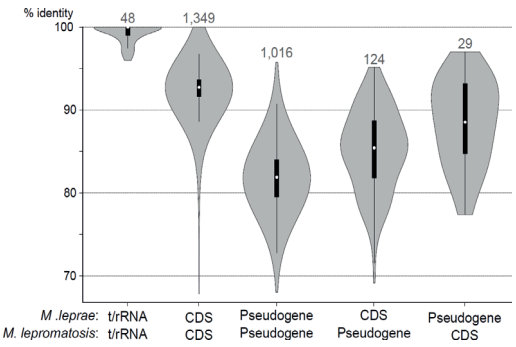


Figure 2: Nucleotide sequence identity between *M. leprae* and *M. lepromatosis* orthologs. Numbers above violins represent the number of ortholog pairs for that gene category. The most conserved are tRNA and rRNA genes, followed by the CDS. Pseudogenes are the least conserved. Identity of genes that are functional in one species but not in the other is between CDS and pseudogenes.

these is present in *M. lepromatosis* (*MLPM_1304*). Likewise, a duplicated cluster of four genes in *M. leprae* (*ML1053-ML1056*; *ML1180c-ML1183c*) is present as a single copy in *M. lepromatosis* (*MLPM_1053MLPM_1056*). This cluster encodes a member of the PEand PPEprotein families and the ESAT-6 proteins, EsxL and EsxK. *M. lepromatosis* has intact orthologs for 95% of the CDS present in *M. leprae*, but a further 132 CDS appear to have been pseudogenized in *M. lepromatosis* (Table S2). It is noteworthy that four of them (*ilvX*, *proA*, *cysE*, *cysK*) once encoded enzymes required for amino acid biosynthesis. Twenty-six *M. leprae* pseudogenes appear to have functional counterparts in *M. lepromatosis* (Table S3). Levels of nucleotide sequence conservation for orthologous CDS and pseudogenes from *M. lepromatosis* and *M. leprae* varied between the functional categories, as may be seen from the violin plot (Fig. 2). The most conserved genes code for the rRNA and tRNA (sequence identity >95%), whereas the least conserved were those for the PE/PPE proteins, as these display the hallmarks of selective pressure (as detailed later). On average, CDS shared 93% nucleotide sequence identity between the two species, but this value was only 82% for the pseudogenes. A wide distribution of sequence conservation was seen among pseudogenes (Fig. 2), and this may reflect their respective dates of pseudogenization, as older pseudogenes will have had longer to diverge.

Species-specific sequences in M. lepromatosis and M. leprae.

We identified 84 genomic regions of *M. lepromatosis* larger than 500 nt that have no counterparts in *M. leprae*, accounting for a total of 166 kb, or ~5% of the genome. These regions (range, 0.5–9.6 kb in size) consist essentially of pseudogenes (n = 163) except for three intact coding sequences: a hypothetical gene (*MLPM_5094*), a putative lipoprotein gene (*MLPM_5098*), and coproporphyrinogen III oxidase (*hemN*) (discussed later). Truncated remnants of 57 of these 163 pseudogenes remain in *M. leprae*, revealing that some of the reductive evolution in this leprosy bacillus stemmed from deletions within, or encompassing, pseudogenes since divergence from the MRCA

Horizontally acquired genes.

Several *M. leprae* genes have no orthologs in other mycobacteria and appear to have been acquired by horizontal gene transfer. Among those with predicted functions are *proS*, encoding a eukaryotic-like prolyl tRNA synthetase, and *ML2177*, coding for a uridine phosphorylase that shows similarity to insect enzymes. Both these genes are conserved in sequence and location in *M. lepromatosis*, indicating that they were present in the MRCA. Likewise, there are two pseudogenes (*MLPM_5100* and *MLPM_5101* encoding a β-lactamase and LysR family transcriptional regulator) in a genomic island restricted to *M. lepromatosis* that appears to have been horizontally acquired by the MRCA and then lost by *M. leprae*.

Insight into pathogenesis.

The higher morbidity and mortality reported to be associated with infection by *M. lepromatosis* and the resulting DLL suggest the presence of new virulence functions possibly borne by plasmids. For example, the pathogenesis of *Mycobacterium ulcerans* has been attributed to the horizontally acquired virulence plasmid encoding the mycolactone toxin (35, 36). However, despite intensive

investigation of Illumina sequence reads with no matches in *M. leprae* and database searches, we found no evidence for plasmid or bacteriophage sequences. To cope with iron limitation, intracellular pathogens often scavenge heme from host tissue or produce and release siderophores, such as mycobactin, to capture iron (37, 38). Both leprosy bacilli have retained the ESX-3 gene cluster that is involved in iron and zinc uptake in *M. tuberculosis*. The mycobactin (*mbt*) gene cluster is essential for the in vivo growth and virulence of *M. tuberculosis* (39, 40), but this cluster is missing from *M. lepromatosis* and *M. leprae.* In *M. tuberculosis*, *hemN* is located downstream of the *mbt* cluster, and, with the *hemABCDEKLYZ* genes, is required for heme biosynthesis. Interestingly, the *hemN* gene is present in *M. lepromatosis* but not in *M. leprae* (Fig. 3), indicating its loss occurred after separation from the MRCA and suggesting that *M. leprae* may be limited for heme production.

Inspection of the genes least conserved between *M. lepromatosis* and *M. leprae* (Fig. 2) revealed that these correspond mainly to members of the PE and PPE protein families, characteristic of pathogenic mycobacteria, and to the ESX (type 7) protein secretion systems. Compared with *M. tuberculosis* and *Mycobacterium marinum*, there are very few PE and PPE proteins in leprosy bacilli. In *M. leprae*, *ML0411* encodes a PPE protein that acts as an immunodominant serine-rich antigen, whereas the neighboring gene, *ML0410*, codes for a PE family protein. Both genes are present in *M. lepromatosis* but have di verged extensively: there is only 68% nucleotide sequence identity between *MLPM_0411* and *ML0411* (note that only half the genes could be aligned) and 73% between *MLPM_0410* and *ML0410*. Interestingly, sequence comparison of *ML0411* in a range of isolates of *M. leprae* of different geographical origin revealed this to be the most polymorphic gene in the genome, with the highest number of nonsynonymous substitutions (20, 41). It appears that the marked divergence in this locus is the consequence of selective pressure imparted by the host's immune system (20).

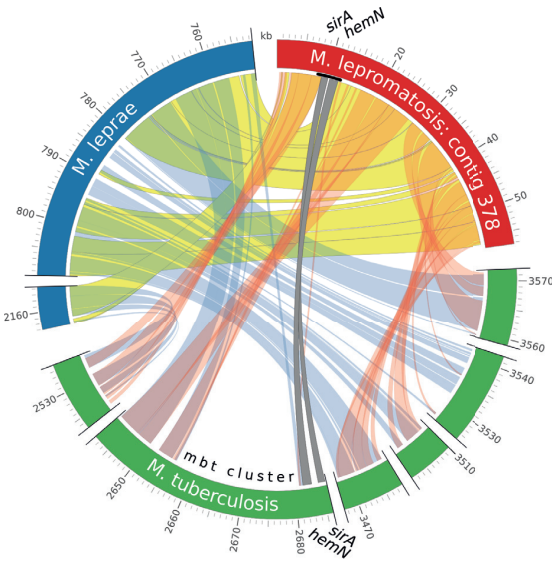


Figure 3: Synteny around the *M. lepromatosis hemN* locus and comparison with *M. leprae* and *M. tuberculosis*. Links between the genomes are BLAST hits; *hemN* and *sirA* are present in *M. lepromatosis* and *M. tuberculosis* (gray links) but have been deleted in *M. leprae*. The *mbt* cluster, present in *M. tuberculosis*, is deleted in *M. leprae* and *M. lepromatosis*. Scale is in kilobases. Note the genomic rearrangement between *M. lepromatosis* (first 10 kb of the contig) and *M. leprae*. Additionally, the sequence corresponding to the *M. leprae* genome between 780 and 795 kb is almost entirely deleted in *M. lepromatosis*

The ability to invade the endothelium distinguishes *M. lepromatosis* from *M. leprae.* Because we found no evidence for the presence of a novel virulence gene that could account for this phenotype, we examined the gene clusters required to produce the five type 7 ESX secretion systems (T7Ss) for unusual features. Although not essential for growth *in vitro*, the ESX-1 system is the major virulence determinant in *M. tuberculosis* and *M. marinum.* ESX-1 mediates escape of the bacilli from the phagosome, thus allowing further replication, cytolysis, necrosis, and intercellular spread (42). The EsxA protein, a major substrate of ESX-1, ruptures the phagosomal membrane, thus acting as a principal virulence factor. Despite extensive conservation of the ESX-1 system among other mycobacterial pathogens, several of its genes encoding ESX-1 secreted proteins (Esp, Esx) are missing or nonfunctional in *M. leprae* and *M. lepromatosis*. These include *espE*, *espB*,

espF, *espG1*, *espH*, *espJ*, *espK*, and *pe35* in both, and, additionally, *ppe68* in *M. lepromatosis*. Of the remaining genes in this locus, *esxB*, *esxA*, and *espH* are the least conserved (69–73% protein identity) between *M. leprae* and *M. lepromatosis*, and a similar trend was observed for the unlinked *espACD* operon (*espA*, 78%; *espC*, 77%; and *espD*, 86% protein identity) that regulates the expression and secretion of EsxA in a mutually dependent manner in *M. tuberculosis* (43–45). Of the four other T7Ss, ESX-2 and ESX-4 are predicted to be nonfunctional in *M. leprae* and *M. lepromatosis*, whereas the ESX-5 locus is highly conserved. However, compared with *M. tuberculosis*, only the ESX-5 core genes (*eccABCDE*$_5$ and *mycP5*) remain intact in both leprosy bacilli, there are no *esx* genes, and the PE protein gene (*ML2534c*/*MLPM_2534*) is nonfunctional. On the contrary, ESX-3 is the most conserved T7S system in mycobacteria and seems to fulfill an essential function in metal homeostasis, although its role in virulence is less clear (45).

Neuropathogenesis.

The ability to invade the Schwann cells of the peripheral nervous system is a hallmark of *M. leprae*, and this leads to the neuropathy and nerve damage associated with leprosy. Adherence to Schwann cells (46) has been proposed to be mediated by two cell wall components: the laminin-binding protein (ML1683c) and the terminal trisaccharide moiety of phenolic glycolipid 1 (PGL-1). To produce the trisaccharide, several enzymes are required, namely a rhamnosyl transferase (ML0128), a glucosyltransferase (ML2348), and four methyltransferases (ML0126, ML0127, ML23246c, and ML2347) (47). The genes encoding both these adhesin systems are highly conserved in *M. lepromatosis*, so invasion of Schwann cells is to be expected. Given the paucity of well-defined cases of infection with *M. lepromatosis*, studies of nerve involvement have not yet been conducted, but this will be facilitated by the tools arising from our investigation of the *M. lepromatosis* genome.

Disease management and new interventions.

Until very recently, recognition of *M. lepromatosis* as a separate species was questioned. Currently, *M. lepromatosis* can be identified by a nested PCR technique that targets the 16S rDNA (22). Given the 98% identity of the *M. lepromatosis* and *M. leprae* 16S rDNA sequences, a possible source of confusion, it may be advisable to establish a new method of identification that exploits sequences confined to *M. lepromatosis* such as the species-specific PCR templates and primers described here (Table 1 and Table S4). Immunodiagnostic approaches for leprosy are being pursued by using the highly specific trisaccharide from PGL-1 and LID1 (48), a fusion protein that includes sequences from genes *ML0405* (*espA*) and *ML2331*. PGL-1-based tests should also detect *M. lepromatosis*, but tests involving LID-1 may be less sensitive because of the extensive variation in EspA reported earlier. DLL cases resulting from infection with *M. lepromatosis* have responded favorably to the standard MDT (49), and, on examination of the drug resistance determining regions in the genes coding for the targets of rifampin, dapsone, and fluoroquinolones, only drug-susceptible sequences were found. New drugs developed to treat tuberculosis may also find application in leprosy. Inspection of the *M. lepromatosis* and *M. leprae* gene sequences for the targets of the experimental drugs bedaquiline, benzothiazinones, and Q203 suggest that these should be active, whereas the nitroimidazole prodrugs PA-824, TBA354, and delamanid will not be effective because the *ddn* gene, coding for the nitroreductase required for their activation, is missing.

Geographical survey for M. lepromatosis.

To gain more insight into the global distribution of *M. lepromatosis*, a differential PCR test was implemented by using species-specific primers targeting *hemN* and RLEP (*SI Materials and Methods*). A total of 227 specimens from patients with leprosy were chosen (Table 1), including some with a history or suspicion of DLL or Lucio's reaction (n = 6). The largest patient groups were from Venezuela (n = 77), Mexico (n = 64), Mali (n = 48), and Brazil (n = 33); a

small number of samples suspected of presenting with Lucio's reaction were obtained from elsewhere (Table 1). This analysis revealed the presence of *M. leprae* DNA in 221 cases and *M. lepromatosis* in only six, with no evidence for mixed infections. All six *M. lepromatosis* cases were of Mexican origin.

Table 1: Geographical survey of leprosy bacilli by differential PCR analysis

Country of origin	Sample size (suspicion of DLL /Lucio's)	*M. lepromatosis*	*M. leprae*
Venezuela	77	0	77
Mexico	64 (4)	6	58
Mali	48	0	48
Brazil	33	0	33
Others	5 (2)	0	5

PCR was performed using primers specific for each species: LPM244F and LPM44R for *M. lepromatosis*; RLEP-7 and RLEP-8 for *M. leprae*. Full details are provided in *SI Materials and Method.* Among the samples from DLL or Lucio's reaction cases, two were positive for *M. lepromatosis* and four for *M. leprae*

On the origin and evolution of M. lepromatosis.

The results from the genome-wide comparison of *M. leprae* and *M. lepromatosis* indicate that pseudogenization took place in their ancestral forms and that the MRCA itself likely had a genome of reduced size compared with all other known mycobacteria. Based on the rate of nonsynonymous substitution in pseudogenes, it was estimated that a single massive pseudogenization event took place approximately 20 Mya (50). Since their separation from the MRCA, deletions and more pseudogenes have appeared in both species. *M. lepromatosis* shares the same repeat families with *M. leprae*, including their genomic locations, but these have diverged extensively in sequence. In *M. leprae*, genome reduction and gene truncation have been attributed to recombinational events between different repeat copies, but these events likely occurred in the MRCA (17, 33).

To estimate the divergence times of *M. lepromatosis* and the currently available *M. leprae* strains, substitution rates were calculated in a Bayesian framework by using the software package BEAST (Bayesian Evolutionary Analysis Sampling Trees). For phylogenetic and divergence time analysis, the *M. lepromatosis* genome was aligned with 18 modern and ancient *M. leprae* genomes. On average, there were 90 nucleotide substitutions between two *M. leprae* strains and 275,518 substitutions between *M. leprae* and *M. lepromatosis*. A substitution rate of 7.67^{-9} substitutions per site per year was estimated (Fig. S2*A*), similar to previous estimates using *M. leprae* genomes only (20). The resulting divergence time from the MRCA (TMRCA) for all *M. leprae* strains was calculated as 3,607 y ago [2,204–5,525 y ago 95% highest probability density (HPD)], comparable to previous results (20) (Fig. S2*B*). The TMRCA for *M. leprae* and *M. lepromatosis* was estimated to be 13.9 Mya (8.2–21.4 Mya 95% HPD; Fig. S2*C*). In this respect, the two leprosy bacilli differ quite markedly from the species comprising the *M. tuberculosis* complex (MTBC), as, from a recent paleomicrobiogical investigation that

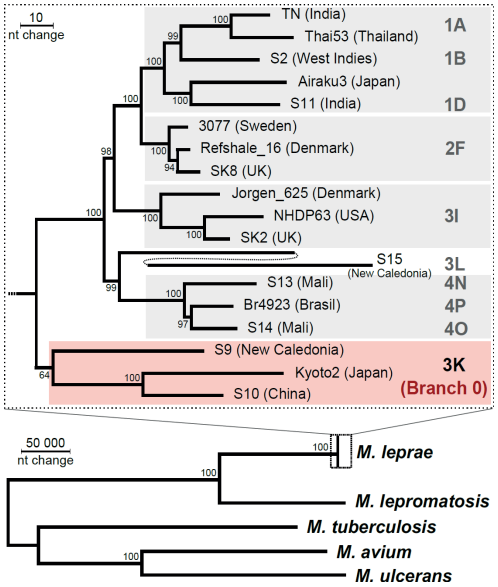


Figure 4: Phylogeny of *M. lepromatosis* and *M. leprae* strains. Phylogenetic relationship of *M. lepromatosis* genome using a maximum parsimony tree, including *M. tuberculosis, M. avium*, and *M. ulcerans* as outgroups. SNP type is given at branch tips. Bootstrap support is indicated for each node. The long branch of S15 was split to reduce space

used two independent dating approaches, it

was concluded that the maximal TMRCA was <6,000 y for the MTBC (51).

A phylogenetic comparison by maximum parsimony of *M. lepromatosis* with various *M. leprae* strains of different geographic origins and SNP subtypes is shown in Fig. 4 and by neighbor joining in Fig. S3. To gain further insight into the phylogenetic placement of *M. lepromatosis* among other mycobacterial species, additional phylogenetic trees were generated based on the concatenated amino acid alignments of GyrB, RpoB, and RpoC (Fig. S4), as well as the alignments of the 16S genes (Fig. S5). In all four phylogenetic trees, *M. lepromatosis* is positioned between the mycobacteria used as outgroups and *M. leprae*. The *M. lepromatosis* branch is closest to the *M. leprae* SNPtype 3K strains, consistent with the very recent report of the 3K strains (branch 0) being the most ancestral lineage of *M. leprae* known to date (20). The geographic distribution of type 3K strains is diverse, but sampling is insufficiently broad to predict a tentative origin. On the contrary, the predominance of reported DLL and confirmed cases of infection with *M. lepromatosis* in Mexico suggest that this pathogen may have evolved in Central America. More extensive investigation is required to explore this possibility and to retrace its origin.

Environmental Sources.

M. leprae has long been considered an obligate human pathogen with its primary route of transmission being interhuman. In the past 30 y, the nine-banded armadillo has been identified as a natural reservoir of *M. leprae* in the southern United States, and evidence for zoonotic transmission to humans is accumulating (34). The existence of nonhuman hosts or natural reservoirs of *M. lepromatosis* has not yet been investigated, but, recently, an apparent outbreak of *M. lepromatosis* infection in red squirrels (*Sciurus vulgaris*) was reported in the United Kingdom (52). Examination of infected squirrel tissue using the PCR-based sequencing procedures for the *hemN* gene outlined here has confirmed this finding and thus provides evidence for a nonhuman reservoir or intermediate host for *M. lepromatosis.*

To date, the most convincing human cases of *M. lepromatosis* infection, particularly those with DLL, came from Mexico. One of the regional culinary traditions in rural Mexico is consumption of field rats (*Rattus rattus*) (53). Because *M. lepromatosis* seems to naturally infect rodents, such as squirrels, it is conceivable that field rats are a host and may serve as a disease reservoir. When a retrospective survey of confirmed cases of *M. lepromatosis* was conducted, two of the six Mexican subjects admitted having consumed meat from field rats. Availability of the *M. lepromatosis* genome sequence will enable us to search systematically for its presence in other potential animal reservoirs as well as in extant cases of leprosy. As a result, deeper understanding will be obtained of the incidence, etiology, epidemiology, clinical features, and pathogenesis of leprosy caused by *M. lepromatosis*.

Materials and Methods

DNA Extraction and Sequencing.

DNA was extracted from the biopsy specimen Mx1-22A (32), and the Illumina library was prepared as described elsewhere (20). To overcome the problem of the high level of host DNA in the library, we used two methods for enrichment of *M. lepromatosis* DNA: (*i*) wholegenome array capture using an *M. leprae* tiling array (20) and (*ii*) removal of human DNA by hybridization with a human genomic DNA bait library.

Genome Assembly and Annotation.

Reads mapping to the human genome were discarded and duplicate reads were removed before the assembly. De novo genome assembly was done in MIRA version 4.0rc4 (54). Contigs were anchored to the *M. leprae* TN genome, and those contigs that did not match were screened for contaminants and the presence of mycobacterial genes. Genome sequence synteny between *M. lepromatosis* and *M. leprae* was visualized in ACT (Artemis Comparison Tool) (55) (Fig. 1).

PCR Procedures.

To confirm the presence of the *hemN* gene and its flanking sequence, including a genomic rearrangement, we designed PCR primers (Table S4) to amplify overlapping genomic targets. Details of the procedures used for the geographical survey are provided in *SI Materials and Methods*.

SNP Calling and Phylogeny. Genome sequences of *Mycobacterium avium* K10

M. lepromatosis, and the published *M. leprae* strains were aligned against the *M. leprae* TN genome. SNP alignments were analyzed in MEGA6 (Molecular Evolutionary Genetics Analysis version 6.0) (56). Further details of bioinformatics procedures, Bayesian dating analysis, and phylogeny are provided in *SI Materials and Methods*.

Acknowledgments: We thank Keith Harshman, Johann Weber, Richard Truman, Rahul Sharma, and Masanori Kai for helpful discussions and technical assistance. This work was supported by grants from the Fondation Raoul Follereau and the Swiss National Science Foundation (Brazilian Swiss Joint Research Program).

References

1. World Health Organization (2013) Global leprosy: Update on the 2012 situation. *Wkly Epidemiol Rec* 88(35):365–379.

2. Ridley DS, Jopling WH (1966) Classification of leprosy according to immunity. A fivegroup system. *Int J Lepr Other Mycobact Dis* 34(3):255–273.

3. Britton WJ, Lockwood DN (2004) Leprosy. *Lancet* 363(9416):1209–1219.

4. Scollard DM, et al. (2006) The continuing challenges of leprosy. *Clin Microbiol Rev* 19(2):338–381.

5. Lucio R, Alsarado Y (1852) *Opusculo sobre el mal de San Làzaro, ó elefanciasis de los griegos, escrito por los profesores de medicina y cirugia* (Imprento de M. Murguia y Compañia, Mexico City, Mexico).

6. Latapi F, Chevez-Zamora A (1948) The "spotted" leprosy of Lucio: An introduction to its clinical and histological study. *Int J Lepr* 16:421–437.

7. Rea TH (1979) Lucio's phenomenon: An overview. *Lepr Rev* 50(2):107–112.

8. Rea TH, Jerskey RS (2005) Clinical and histologic variations among thirty patients with Lucio's phenomenon and pure and primitive diffuse lepromatosis (Latapi's lepromatosis). *Int J Lepr Other Mycobact Dis* 73(3):169–188.

9. Moschella SL (1967) The lepra reaction with necrotizing skin lesions. A report of six cases. *Arch Dermatol* 95(6):565–575.

10. Romero A, Brenes, Ibarra A, Fallas M (1949) Clinical study of lepromatous leprosy in Costa Rica. *Int J Lepr* 17(1-2):27–33.

11. Vargas-Ocampo F (2007) Diffuse leprosy of Lucio and Latapí: A histologic study. *Lepr Rev* 78(3):248–260.

12. Sehgal VN (2005) Lucio's phenomenon/erythema necroticans. *Int J Dermatol* 44(7): 602–605.

13. Ang P, Tay Y-K, Ng S-K, Seow C-S (2003) Fatal Lucio's phenomenon in 2 patients with previously undiagnosed

leprosy. *J Am Acad Dermatol* 48(6):958–961.

14. Peixoto AB, Portela PS, Leal FRP de C, Brotas AM, Rodrigues NC dos S (2013) Lucio's phenomenon. Case study of an exceptional response to treatment exclusively with multibacillary multidrug therapy. *An Bras Dermatol* 88(6, Suppl 1):93–96.

15. World Health Organization (2014) WHO Recommended MDT Regimens. Available at www.who.int/lep/mdt/regimens/en/. Accessed September 24, 2014.

16. Helmer KA, Fleischfresser I, Kucharski-Esmanhoto LD, Fillus Neto J, Santamaria JR (2004) The Lucio's phenomenon (necrotizing erythema) in pregnancy. *An Bras Dermatol* 79(2):205–210.

17. Cole ST, et al. (2001) Massive gene decay in the leprosy bacillus. *Nature* 409(6823): 1007–1011.

18. Williams DL, Gillis TP, Portaels F (1990) Geographically distinct isolates of Mycobacterium leprae exhibit no genotypic diversity by restriction fragment-length polymorphism analysis. *Mol Microbiol* 4(10):1653–1659.

19. Monot M, et al. (2009) Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nat Genet* 41(12):1282–1289.

20. Schuenemann VJ, et al. (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341(6142):179–183.

21. Han XY, et al. (2008) A new Mycobacterium species causing diffuse lepromatous leprosy. *Am J Clin Pathol* 130(6):856–864.

22. Han XY, Sizer KC, Velarde-Félix JS, Frias-Castro LO, Vargas-Ocampo F (2012) The leprosy agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico. *Int J Dermatol* 51(8):952–959.

23. Han XY, Sizer KC, Tan H-H (2012) Identification of the leprosy agent *Mycobacterium lepromatosis* in Singapore. *J Drugs Dermatol* 11(2):168–172.

24. Jessamine PG, et al. (2012) Leprosy-like illness in a patient with Mycobacterium lepromatosis from Ontario, Canada. *J Drugs Dermatol* 11(2):229–233.

25. Monteiro R, et al. (2012) Lucio's phenomenon: Another case reported in Brazil. *An Bras Dermatol* 87(2):296–300.

26. Souza CS, Roselino AM, Figueiredo F, Foss NT (2000) Lucio's phenomenon: Clinical and therapeutic aspects. *Int J Lepr Other Mycobact Dis* 68(4):417–425.

27. Saoji V, Salodkar A (2001) Lucio leprosy with lucio phenomenon. *Indian J Lepr* 73(3): 267–272.

28. Thappa DM, Karthikeyan K, Kumar BJ (2002) Is it Lucio leprosy with Lucio phenomenon or something else? *Indian J Lepr* 74(2):161–166.

29. Kumari R, Thappa DM, Basu D (2008) A fatal case of Lucio phenomenon from India. *Dermatol Online J* 14(2):10.

30. Golchai J, Zargari O, Maboodi A, Maboodi A, Granmayeh S (2004)

Lepromatous leprosy with extensive unusual ulcerations and cachexia. Is it the first case of Lucio's phenomenon from Iran? *Int J Lepr Other Mycobact Dis* 72(1):56–59.

31. Choon SE, Tey KE (2009) Lucio's phenomenon: A report of three cases seen in Johor, Malaysia. *Int J Dermatol* 48(9):984–988.

32. Vera-Cabrera L, et al. (2011) Case of diffuse lepromatous leprosy associated with

33. "*Mycobacterium lepromatosis*". *J Clin Microbiol* 49(12):4366–4368.

34. Cole ST, Supply P, Honoré N (2001) Repetitive sequences in *Mycobacterium leprae* and their impact on genome plasticity. *Lepr Rev* 72(4):449–461.

35. Truman RW, et al. (2011) Probable zoonotic leprosy in the southern United States. *N Engl J Med* 364(17):1626–1633.

36. Stinear TP, et al. (2007) Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. *Genome Res* 17(2):192–200.

37. Demangel C, Stinear TP, Cole ST (2009) Buruli ulcer: Reductive evolution enhances pathogenicity of Mycobacterium ulcerans. *Nat Rev Microbiol* 7(1):50–60.

38. Chu BC, et al. (2010) Siderophore uptake in bacteria and the battle for iron with the host; a bird's eye view. *Biometals* 23(4):601–611.

39. Braun V, Hantke K (2011) Recent insights into iron import by bacteria. *Curr Opin Chem Biol* 15(2):328–334.

40. De Voss JJ, et al. (2000) The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages. *Proc Natl Acad Sci USA* 97(3):1252–1257.

41. Reddy PV, et al. (2013) Disruption of mycobactin biosynthesis leads to attenuation of *Mycobacterium tuberculosis* for growth and virulence. *J Infect Dis* 208(8):1255–1265.

42. Kai M, et al. (2013) Characteristic mutations found in the ML0411 gene of Mycobacterium leprae isolated in Northeast Asian countries. *Infect Genet Evol* 19:200–204.

43. Simeone R, et al. (2012) Phagosomal rupture by *Mycobacterium tuberculosis* results in toxicity and host cell death. *PLoS Pathog* 8(2):e1002507.

44. Fortune SM, et al. (2005) Mutually dependent secretion of proteins required for mycobacterial virulence. *Proc Natl Acad Sci USA* 102(30):10676–10681.

45. Garces A, et al. (2010) EspA acts as a critical mediator of ESX1-dependent virulence in *Mycobacterium tuberculosis* by affecting bacterial cell wall integrity. *PLoS Pathog* 6(6):e1000957.

46. Stoop EJM, et al. (2011) Zebrafish embryo screen for mycobacterial genes involved in the initiation of granuloma formation reveals a newly identified ESX-1 component. *Dis Model Mech* 4(4):526–536.

47. Rambukkana A (2001) Molecular basis for the peripheral nerve predilection of Mycobacterium leprae. *Curr Opin Microbiol* 4(1):21–27.

48. Tabouret G, et al. (2010) *Mycobacterium leprae* phenolglycolipid-1 expressed by engineered M. bovis BCG modulates early interaction with human phagocytes. *PLoS Pathog* 6(10):e1001159.

49. de Souza MM, Netto EM, Nakatani M, Duthie MS (2014) Utility of recombinant proteins LID-1 and PADL in screening for *Mycobacterium leprae* infection and leprosy. *Trans R Soc Trop Med Hyg* 108(8):495–501.

50. Han XY, Jessurun J (2013) Severe leprosy reactions due to Mycobacterium lepromatosis. *Am J Med Sci* 345(1):65–69.

51. Han XY, et al. (2009) Comparative sequence analysis of *Mycobacterium leprae* and the new leprosy-causing *Mycobacterium lepromatosis*. *J Bacteriol* 191(19):6067–6074.

52. Bos KI, et al. (2014) Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature* 514(7523):494–497.

53. Meredith A, et al. (2014) Leprosy in red squirrels in Scotland. *Vet Rec* 175(11):285–286.

54. Chacon R (October 8, 1999) In Mexico, field rats are secret delicacy. *Miami Herald*. Available at www.latinamericanstudies.org/mexico/rats.htm. Accessed October 16, 2014.

55. Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, Vol 99, pp 45–56. Available at www.bioinfo.de/isb/gcb99/talks/chevreux/main.html.

56. Carver TJ, et al. (2005) ACT: The Artemis Comparison Tool. *Bioinformatics* 21(16): 3422–3423.

57. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729.

Supplementary materials and methods

Clinical Specimens

The original skin biopsy specimen from an 86-yold Mexican woman diagnosed with DLL (Mx1-22A) was shipped to Ecole Polytechnique Fédérale de Lausanne (Switzerland) in 70% (vol/vol) ethanol for DNA extraction and sequencing. A detailed case report for this sample was published previously (1). All other biopsies were from our collections.

DNA Extraction and Sequencing

DNA was extracted from the biopsy specimen Mx1-22A (1) and the Illumina library (labeled mmL1) was prepared as described elsewhere (2). Aliquots of the mmL1 library were used for two additional enrichment procedures: whole-genome capture array and removal of human DNA by hybridization. Because the *Mycobacterium lepromatosis* genome sequence exhibits high homology with that of *Mycobacterium leprae*, we used a custom-synthesized oligonucleotide array (Agilent) spanning the entire *M. leprae* genome (2) to capture homologous sequences in *M. lepromatosis*. This array-enriched library was labeled mmL1-p (positive selection). However, as this strategy would miss those genomic regions of *M. lepromatosis* that have weak or no homology with *M. leprae*, we also used an enrichment strategy to deplete human DNA. This involved preparing a bait library with human genomic DNA by using the Illumina TruSeq kit followed by amplification with the following biotinylated primers.

Primer	Sequence 5'-3'	Purpose
P5	AATGATACGGCG ACCACCGA	Amplification of target library
Bio-P5	5' Biotin-AATGATACGGCG ACCACCGA	Amplification of human-bait library
P7	CAAGCAGAAGAC GGCATACGA	Amplification of human-bait and target library

The resultant biotinylated baits were denatured and coated on streptavidin magnetic beads to prepare "bead baits," which were used for the hybridization-capturing of human DNA from the mmL1 library. After removing the bead-bound human DNA, the unbound fraction of the library was amplified and size selected on a 2% (wt/vol) agarose gel. Fragments of 300 bp (library mmL21) and 400 bp (library mmL25) were extracted from the gel and sequenced by using single-end reads on a HiSEq 2000 instrument. The number and the properties of the reads obtained from each of the four libraries are shown in Table S1. Sequence reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive database under accession number SRP047206.

Genome Assembly and Annotation

Reads were adapter-trimmed with Flexbar (3). Paired-end reads deriving from MiSeq were merged using SeqPrep (https://github.com/jstjohn/SeqPrep). Reads mapping to the human genome (hg19, GRCh37, using Bowtie2 v2.2.2 with "–very-fast-local" parameter), PhiX (NC_001422.1, using Bowtie2 v2.2.2 with "–local" parameter) and *Escherichia coli* (NC_010473, using Bowtie2 v2.2.2 with "–local" parameter) were discarded (4). The remaining reads were first mapped to the *M. leprae* TN genome sequence to assess the *M. lepromatosis* genome coverage and read quality. We noticed a high level of duplicate reads (reads with identical sequence and start-end coordinates), especially in the mmL1-p library (Table S1), which worsened the quality of our preliminary de novo assembly. We removed all duplicate reads by using PRINSEQ (Preprocessing and Information of Sequences) (5) and performed a de novo assembly by using MIRA version 4.0rc4 (6) with the "EST" mode. Contigs were anchored to the *M. leprae* TN genome using BLAST (7), and those with inconsistent synteny with *M. leprae* were split. Contigs that did not align to *M. leprae* were screened for contaminating DNA (human, phiX, and *E. coli*). Contigs consisting entirely of low-complexity sequence were removed. The remaining contigs were

queried with BLAST against mycobacterial proteins from NCBI to detect any putative *M. lepromatosis* contigs that did not anchor to the *M. leprae* genome (one 2.3-kb-long contig was detected). Annotation was carried out by using RATT (Rapid Annotation Transfer Tool) (8) followed by manual curation. *M. leprae* annotation was taken from the MycoBrowser portal (9), release 7. *M. lepromatosis*-specific genomic regions were queried with BLAST against nonredundant proteins at NCBI and manually annotated. Annotated contigs were deposited at DNA Data Bank of Japan/European Molecular Biology Laboratory/GenBank under accession number JRPY00000000.

PCR and Sanger sequencing

To confirm the presence of the *hemN* gene and its flanking sequence, including a genomic rearrangement in *M. lepromatosis*, we designed PCR primers (Table S4) to amplify overlapping genomic targets. PCR products were purified with the Illustra ExoProStar 1-Step cleanup kit (GE Healthcare) and sequenced on an ABI PRISM 3130xl Genomic Analyzer (Applied Biosystems). For the geographic survey of leprosy bacilli, PCR amplification was performed by using the following primers.

Primer name	Primer sequence (5'-3')
LPM244-F	GTTCCTCCACCGACAAACAC
LPM244-R	TTCGTGAGGTACCGGTGAAA
RLEP-7	TGAGGCTTCGTGTGCTTTGC
RLEP-8	ATCTGCGCTAGAAGGTTGCC

LPM244 primers amplify a 244-bp fragment from the *hemN* gene of *M. lepromatosis* that has been lost by *M. leprae*, whereas RLEP 7 and 8 (10) amplify the RLEP repetitive sequences in *M. leprae*. PCR amplifications were performed in a 50-μL reaction volume containing 25 μL of AccuStart II SuperMix (Quanta BioSciences) or BIO-X-ACT Short Mix (Bioline), 5 μL of each primer at 2 μM, and 2 μL of DNA extracts or sterile distilled water for clinical specimens, positive controls (Mx1-22 and Thai-53 strains), and negative controls. The amplification started with an initial denaturation step at 95 °C followed by 40 cycles of 30 s denaturation at 95 °C, 40 s primer annealing at 58 °C, and 30 s extension at 72 °C, followed by a 10-min final extension at 72 °C. The amplicons obtained were then analyzed by electrophoresis on a 1% agarose gel in 1× Tris/acetate/EDTA buffer at 100 V during 35 min. The DNA was stained using GelRed (Chemie Brunschwig) and visualized by using a UV transilluminator. After enzymatic treatment with Illustra ExoProStar 1-Step (GE Healthcare Life Sciences), PCR products were submitted to BigDye Terminator version 3.1 cycle sequencing and analyzed by using an ABI3130 XL DNA sequencer (Applied Biosystems). Sequences were aligned against corresponding references using CodonCode aligner software (version 5.0.1; CodonCode).

SNP calling.

M. lepromatosis contigs, *Mycobacterium avium* K10 (a.n. NC_002944), *M. tuberculosis* H37Rv (NC_ 000962.3), and *Mycobacterium ulcerans* Agy99 (NC_ 008611.1), and the published genomes of *M. leprae* (10, 11) were aligned against *M. leprae* TN (accession no. AL450380.1) with LAST (12) using the gammacentroid mode for *M. lepromatosis*, *M. avium*, *M. tuberculosis*, and *M. ulcerans* and default parameters for *M. leprae* strains. For each alignment, an mpileup file was generated with SAMtools version 0.1.19 (13) and processed with VarScan version 2.3.7 (14). For *M. lepromatosis*, sites with >1× coverage (repetitive areas) were removed, whereas, for *M. avium*, *M. tuberculosis*, and *M. ulcerans*, sites with >1× coverage (gene duplications) were kept except for heterozygous sites. Illumina reads from the *M. leprae* strains published in the work of Schuenemann et al. were aligned as described there (2), and the Illumina reads from Airaku-3 (15) were aligned against TN with Bowtie2 version 2.2.2 (4). Mpileup files were generated and processed with VarScan (14). Minimum coverage was set to 3, and minimum SNP frequency was set to 80%. Indels were filtered out, as were the SNPs with frequencies 20–80%. Positions covered in the negative control SK12 (2) were removed.

Estimation of substitution rates and divergence times.

To estimate the divergence times of the *M. lepromatosis* and *M. leprae* strains, substitution rates were calculated in a Bayesian framework by using the software package BEAST 1.7.5 (16). A multiple sequence alignment of variable positions between the reconstructed *M. lepromatosis* genome (LPM) as well as previously published ancient *M. leprae* genomes that were directly radiocarbon-dated (Jorgen_625, Refshale_16, 3077, SK8, SK2) and available modern samples (Kyoto2, Airaku3, TN, Br4923, Thai53, NHDP63, S2, S9, S10, S11, S13, S14), was made as described earier. The total alignment length was 719,495 variable positions. As described previously, the *M. leprae* genome S15 was excluded from the dating analysis because of the rather long branch observed for that strain, which might be a result of selection pressure caused by antileprosy treatment as reflected by the occurrence of a dapsone-resistance mutation (17). Based on the previous analysis, a strict clock model with constant population size was chosen for the Bayesian framework analysis. The previously published calibrated radiocarbon dates for the five ancient genomes, as well as the isolation years for the modern genomes, were used as priors (2). Three Markov chain Monte Carlo (MCMC) runs were carried out with 50,000,000 iterations each, sampling every 10,000 steps. The first 5,000,000 iterations were discarded as burn-in. All independent runs were combined, resulting in 135,000,000 iterations.

We estimate a substitution rate of 7.67^{-9} substitutions per site per year (1.11×10^{-8} to 4.2×10^{-9} 95% HPD; Fig. S2*A*), which is very similar to a previous estimate using *M. leprae* genomes only (2). The resulting divergence times for the TMRCA for all *M. leprae* strains is 3,607 y ago (2,204–5,525 y ago 95% HPD; Fig. S2*B*), and the TMRCA for *M. leprae* and *M. lepromatosis* is 13.9 Mya (8.2–21.4 Mya 95% HPD; Fig. S2*C*).

Phylogeny

A multiple-sequence alignment of variable positions between the reconstructed *M. lepromatosis* genome, previously published *M. leprae* genomes, *M. tuberculosis*, *M. avium*, and *M. ulcerans* had a total of 1,033,823 sites. All trees (Fig. 4 and Figs. S3–S5) were generated in MEGA6 (18) using 500 bootstrap replicates. For the maximum parsimony (MP) tree in Fig. 4, sites with missing data were partially deleted (90% coverage cutoff), resulting in 899,085 variable sites used for the tree calculation. Subtree-pruning-regrafting was used as the MP search method. For the neighbor joining tree (Fig. S3), sites with missing data were completely removed, resulting in 511,280 variable sites. The number of nucleotide differences was used as model, with uniform rates among sites. As expected, both trees strongly resemble each other.

To gain further insight into the phylogenetic placement of *M. lepromatosis* among other mycobacterial species, additional phylogenetic trees were generated based on the concatenated amino acid alignments of GyrB, RpoB, and RpoC (Fig. S4), as well as the alignments of the 16S rRNA genes (Fig. S5). Protein sequences were downloaded from the HOGENOM database (19) and aligned with Clustal Omega (20). To reduce the size of the tree, only representative species that branched around *M. lepromatosis* were selected for the generation of the final tree, with *Rhodococcus equi* as outgroup. The maximum-likelihood tree was generated using the LG model with frequencies (21) with rates among sites as "Gamma distributed with Invariant sites" (G+I), as this model had the lowest Bayesian information criterion score for this dataset as calculated by MEGA6, and missing data were partially deleted (90% coverage cutoff). 16S rRNA sequences were downloaded from the Greengenes database (22) and aligned with CRWalign (www.rna.icmb.utexas. edu) (23). For the MP tree (Fig. S5), positions with missing data (i.e., gaps) were completely deleted. To reduce the size of the tree, only

representative species that branched around *M. lepromatosis* were selected for the generation of the final tree, with *Mycobacterium smegmatis* as outgroup. Results from our phylogenetic analysis were consistent with previous studies (24–26).

References

1. Vera-Cabrera L, et al. (2011) Case of diffuse lepromatous leprosy associated with "*Mycobacterium lepromatosis*". *J Clin Microbiol* 49(12):4366–4368.

2. Schuenemann VJ, et al. (2013) Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341(6142):179–183.

3. Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-Flexible Barcode and Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* 1(3):895–905.

4. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

5. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.

6. Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*, Vol 99, pp 45–56. Available at www.bioinfo.de/isb/gcb99/talks/chevreux/main.html.

7. Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.

8. Otto TD, Dillon GP, Degrave WS, Berriman M (2011) RATT: Rapid Annotation Transfer Tool. *Nucleic Acids Res* 39(9):e57.

9. Kapopoulou A, Lew JM, Cole ST (2011) The MycoBrowser portal: A comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)* 91(1):8–13.

10. Monot M, et al. (2009) Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nat Genet* 41(12):1282–1289.

11. Kai M, et al. (2013) Characteristic mutations found in the ML0411 gene of Mycobacterium leprae isolated in Northeast Asian countries. *Infect Genet Evol* 19:200–204.

12. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC (2011) Adaptive seeds tame genomic sequence comparison. *Genome Res* 21(3):487–493.

13. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079

14. Koboldt DC, et al. (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22(3):568–576.

15. Singh P, et al. (2014) Genome-wide re-sequencing of multidrug-resistant Mycobacterium leprae Airaku-3. *Clin Microbiol Infect* 20(10):O619–O622. onlinelibrary.wiley.com/doi/10.1111/1469-0691.12609/abstract.

16. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

17. Maeda S, et al. (2001) Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrob Agents Chemother* 45(12):3635–3639.

18. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729.

19. Penel S, et al. (2009) Databases of homologous gene families for comparative genomics. *BMC Bioinformatics* 10(suppl 6):S3.

20. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.

21. Le SQ, Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* 25(7):1307–1320.

22. DeSantis TZ, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72(7):5069–5072.

23. Shang L, et al. (2013) Two accurate sequence, structure, and phylogenetic templatebased RNA alignment systems. *BMC Syst Biol* 7(suppl 4):S13.

24. Han XY, et al. (2009) Comparative sequence analysis of *Mycobacterium leprae* and the new leprosy-causing *Mycobacterium lepromatosis*. *J Bacteriol* 191(19):6067–6074.

25. Alam MT, Merlo ME, Takano E, Breitling R (2010) Genome-based phylogenetic analysis of Streptomyces and its relatives. *Mol Phylogenet Evol* 54(3):763–772.

26. Gao B, Gupta RS (2012) Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiol Mol Biol Rev* 76(1):66–112.

Table S1. Sequencing libraries of *M. lepromatosis* Mx1-22A.

Library name	Method used for library preparation and enrichment	Platform *	M reads	Mapped onto *M. lpm*	Dupli-cation rate	Fold coverage of *M. lpm* assembly **
mmL1	Library mmL1, from biopsy after silica extraction, no enrichment.	HiSeq PE+SE	372 + 89	0.12%	8%	14.2
mmL1-p	Capture of the library mmL1 on *M. leprae* array.	HiSeq SE MiSeq PE	239 15	3.0% 3.0%	89% 43%	21.6 9.6
mmL21	Removal of human DNA from the library mmL1 using biotinylated Human DNA baits and size selection of ~300 bp fragments.	HiSeq SE	274	0.39%	29%	14.7
mmL25	Removal of human DNA from the library mmL1 using biotinylated Human DNA baits and size selection of ~400 bp fragments.	HiSeq SE	253	0.34%	22%	13.8
All			1,242	0.82%	79%***	54.5***

Legend: M = Million; PE = Paired-end reads, SE = single-end reads; *M. lpm* = *M. lepromatosis*.

*Read length was 100 nt for HiSeq and 150 for MiSeq.

**Excluding duplicate reads.

***Duplicate reads removed after merging the original datasets.

Table S2: ***M. leprae*** **protein-coding genes with disrupted ORFs in** ***M. lepromatosis***

M. leprae[1]	Function
ML0008c*	hypothetical protein
ML0023c*	hypothetical protein
ML0024*	hypothetical protein
ML0029c	hypothetical protein
ML0051c	PPE family protein
ML0064c	transcriptional regulator
ML0188*	hypothetical protein
ML0218c	hypothetical protein
ML0283*	cation-efflux transporter component
ML0293c*	hypothetical protein
ML0308*	hypothetical protein
ML0316	TetR/AcrR family transcriptional regulator
ML0354	acetohydroxyacid synthase IlvX (acetolactate synthase)
ML0363	ESAT-6 like protein EsxT
ML0369c	hypothetical protein
ML0376	hypothetical protein
ML0382c	transcriptional regulator
ML0394c	3-oxoacyl-ACP synthase
ML0398c	D-ribose-binding protein
ML0418	oxidoreductase
ML0447*	hypothetical protein
ML0448*	hypothetical protein
ML0464*	hypothetical protein
ML0470c*	hypothetical protein
ML0508	hypothetical protein
ML0527c	hypothetical protein
ML0580c	OpcA
ML0588*	hypothetical protein
ML0603	hypothetical protein
ML0605	hypothetical protein
ML0606	hypothetical protein
ML0607	hypothetical protein

ML0638	hypothetical protein
ML0639	transcriptional regulator
ML0664	hypothetical protein
ML0676c	hypothetical protein
ML0678c*	hypothetical protein
ML0717c	TetR family transcriptional regulator
ML0728c	thiosulfate sulfurtransferase
ML0729c	Maf-like protein
ML0761c	hypothetical protein
ML0777*	hypothetical protein
ML0799c	hypothetical protein
ML0837	hypothetical protein
ML0838c	serine acetyltransferase, CysE
ML0839c	cysteine synthase, CysK
ML0844	nitrite extrusion protein
ML0845c	hypothetical protein
ML0863	hypothetical protein
ML0923	hypothetical protein
ML0926c*	hypothetical protein
ML0928	hypothetical protein
ML0946	hypothetical protein
ML0953*	hypothetical protein
ML0955c	pyruvate phosphate dikinase
ML0958*	hypothetical protein
ML1006	hypothetical protein
ML1011c*	hypothetical protein
ML1023c*	polyphosphate glucokinase
ML1057	hypothetical protein
ML1067	hypothetical protein
ML1092A	hypothetical protein
ML1103c	oxidoreductase subunit
ML1188c*	hypothetical protein
ML1193	hypothetical protein
ML1294c	hypothetical protein

ML1344c*	hypothetical protein
ML1399	hypothetical protein
ML1420	hypothetical protein
ML1458c	gamma-glutamyl phosphate reductase, ProA
ML1508c	hypothetical protein
ML1525c	hypothetical protein
ML1572*	hypothetical protein
ML1575c*	hypothetical protein
ML1601c*	hypothetical protein
ML1602c	hypothetical protein
ML1603c*	hypothetical protein
ML1632c	exported protease
ML1660c	hypothetical protein
ML1682	hypothetical protein
ML1761c*	hypothetical protein
ML1783c	transcriptional regulator
ML1788c	hypothetical protein
ML1796*	hypothetical protein
ML1818c	thioredoxin
ML1821A*	hypothetical protein
ML1821c*	hypothetical protein
ML1915*	hypothetical protein
ML1937	hypothetical protein
ML1949c*	hypothetical protein
ML1972c*	hypothetical protein
ML1976c*	hypothetical protein
ML1979c*	hypothetical protein
ML1982	hypothetical protein
ML1989c*	hypothetical protein
ML1990c*	hypothetical protein
ML1997c	hypothetical protein
ML2041c	hydrogen peroxide-inducible gene activator OxyR
ML2048*	hypothetical protein
ML2121c*	hypothetical protein

ML2156	hypothetical protein
ML2178c	hypothetical protein
ML2201c*	hypothetical protein
ML2213c	aminopeptidase
ML2242c	hypothetical protein
ML2249c*	hypothetical protein
ML2252c*	hypothetical protein
ML2264c	hypothetical protein
ML2283*	hypothetical protein
ML2284*	hypothetical protein
ML2288c*	hypothetical protein
ML2307c	hypothetical protein
ML2312	hypothetical protein
ML2366c	thiol:disulfide interchange protein
ML2369B*	hypothetical protein
ML2379c	hypothetical protein
ML2388	hypothetical protein
ML2468c*	hypothetical protein
ML2497*	hypothetical protein
ML2564c	acetyl-CoA acetyltransferase
ML2567	hypothetical protein
ML2609	hypothetical protein
ML2621c	hypothetical protein
ML2629c	hypothetical protein
ML2630c	hypothetical protein
ML2661c	acyl-CoA synthetase
ML2666c*	hypothetical protein
ML2679A*	hypothetical protein
ML2679B*	hypothetical protein
ML2696	MarR family transcriptional regulator
ML2703*	thioredoxin
ML2709c	hypothetical protein

[1] Annotation taken from the Mycobrowser database (http://mycobrowser.epfl.ch/leprosy.html). Prefix for *M. lepromatosis* is "MLPM_".

Table S3: *M. leprae* pseudogenes with intact ORFs in *M. lepromatosis*

Pseudogene [1]	Function
ML0062A*	Hypothetical protein
ML0209c	Hypothetical protein
ML0212	Conserved hypothetical protein
ML0302*	Possible conserved membrane protein
ML0358*	Probable ferredoxin FdxD
ML0954c	Conserved hypothetical protein
ML0970A*	Hypothetical protein
ML0976c	Conserved hypothetical protein
ML0981	Conserved 35 kDa ala-rich protein
ML1208	Conserved hypothetical protein
ML1254A*	Hypothetical protein
ML1341	Conserved hypothetical protein
ML1451D*	Hypothetical protein
ML1527c*	Dihydrodipicolinate reductase DapB
ML1531c	Possible transcriptional regulatory protein (Lrp/AsnC family)
ML1636A	Hypothetical protein
ML1643	Probable low molecular weightprotein-tyrosine-phosphatase PtpA
ML1686	Probable transcriptional regulatory protein (IclR family)
ML1721c	Probable conserved secreted protein TB22.2 homolog
ML1725c	Conserved hypothetical protein
ML1745c	Probable acyl-CoA dehydrogenase FadE22
ML2019	Conserved hypothetical protein
ML2119	Putative acetyl-CoA carboxylase carboxyltransferase (beta subunit) AccD3
ML2253A*	Hypothetical protein
ML2476c	Conserved hypothetical protein
ML2633A*	Hypothetical protein

Table S4. Primers used for PCR-sequencing genomic region around *hemN* in *M. lepromatosis*

Forward primer	Sequence	Reverse primer	Sequence
hemN_F1	GGAGGAACCCCAGAGAATGC	hemN_R1	CATCCGGGGGTTTTCCTTCA
hemN_F11	CGGTGCTATGTGGAACAACG	hemN_R11	CTACTGGAACGGCGGTCAAT
hemN_F12	GTTGTGCCGACGTAACTGTG	hemN_R12	TCGGGGTGTATGTGCATGTC
hemN_F13	TATGTGATCGACACCGGCTT	hemN_R13	CCACCTGGCCGGAATTTTTC
hemN_F2	TTGAGTGGCCAACACCGTTA	hemN_R2	TTGATCGCACGCAACACTTG
hemN_F21	AATCTGCGTCGCCAAAACAC	hemN_R21	CCAGTGATTCATCGGCCAGT
hemN_F22	ACGGCAGCTGGACTAATGAC	hemN_R22	CCGGAGTCACGAAACAGTGA
hemN_F23	GGTGATCTGTCGACCAACCC	hemN_R23	GTTCCGCGAAGCCTCTGATA
hemN_F24	TCTCGAAACCGGGTACCTCA	hemN_R24	TCTCGACGTGTTCCTTGGTG
hemN_F25	CTACCCCGATTGGCACACTG	hemN_R25	CCACTGCTCGATCGGATTGT
hemN_F31	TCCGGCACAAGGGTGTAAAA	hemN_R31	GCCGATTGCAGGTCGTTTAG
hemN_F32	GCCACTCCTGCTGTTCTTCT	hemN_R32	CTGGAAAGCCAGGACCAAGA
hemN_F33	TAAGTCAGGCTACACCACGC	hemN_R33	GTTGTTCCACATAGCACCGC
hemN_F34	GTCGCAGGAAACTGTCCTCA	hemN_F34	CTGCTGGGATAACCCGACG
hemN_F41	GGCCAGGTGGATTCAGGATT	hemN_R41	CGAGAATCCGCTGAGAAGCA
hemN_F42	CCGGTACCTCACGAACAGTC	hemN_R42	GTGTGAGCTTTTGCCAGACG
hemN_F43	CTCCTGTCACTGACCCAACC	hemN_R43	CGTCTAGCTTCCGCGAGATT
hemN_F44	AACGGGTTATGCAGAAGGGG	hemN_R44	TTCCCGGGCTTCTTGAGTTC
hemN_F45	ATGTCGCCAACCAGGTATCG	hemN_R42	See above

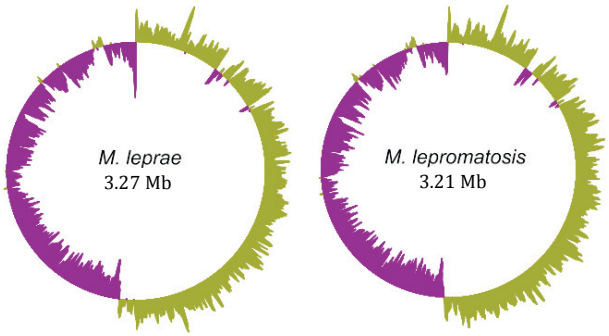


Fig. S1. GC skews of *M. leprae* and *M. lepromatosis* genomes. When *M. lepromatosis* contigs were anchored to the *M. leprae* genome and concatenated, the resulting GC skew was virtually identical to that of *M. leprae*.

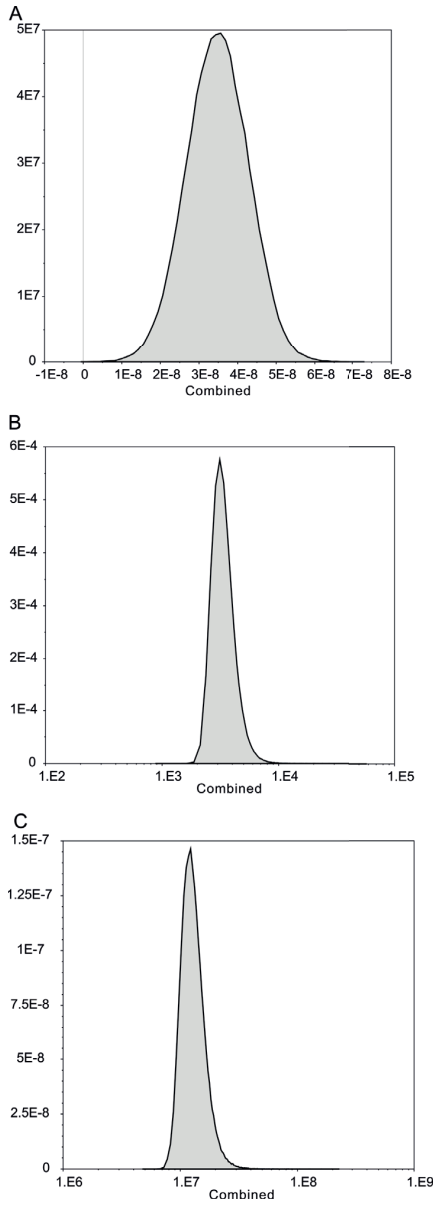


Fig. S2. Bayesian dating analysis of *M. lepromatosis*. (*A*) Substitution rate calculated by tip calibration of radiocarbon dated *M. leprae* genomes using BEAST, a Bayesian framework analysis package. The substitution rate is based on the total alignment of 719,495 variable positions. (*B*) Posterior distribution for the TMRCA for all *M. leprae* strains. (*C*) Posterior distribution for the TMRCA for *M. leprae* strains and *M. lepromatosis*

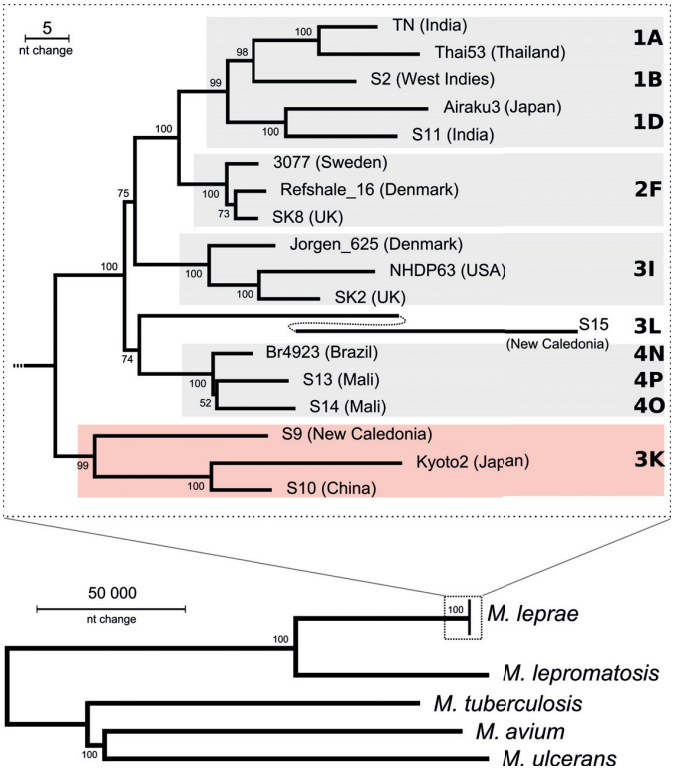


Figure S3: Phylogeny of *M. lepromatosis* and *M. leprae* strains. Phylogenetic relationship of *M. lepromatosis* genome using a neighbor-joining tree, including *M. tuberculosis*, *M. avium*, and *M. ulcerans* as outgroups. SNP type is given at branch tips. Bootstrap support is indicated for each node. The long branch of S15 was split to reduce space

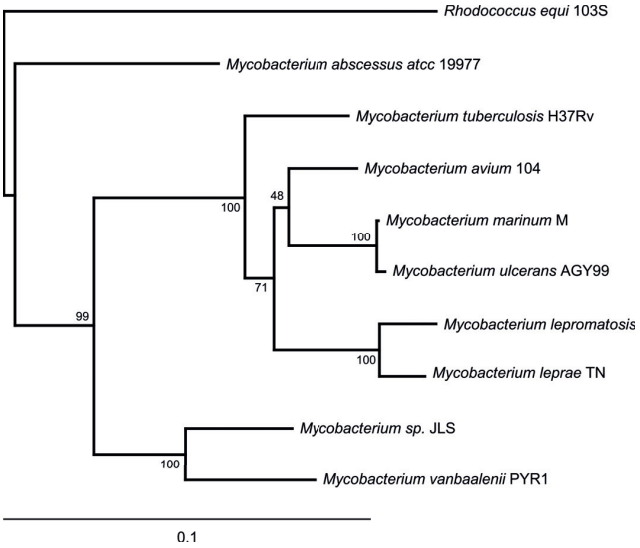


Figure S4: Phylogeny of selected mycobacterial species in a maximum likelihood tree based on the concatenated amino acid sequence alignments of GyrB, RpoB, and RpoC. Bootstrap support is indicated for each node

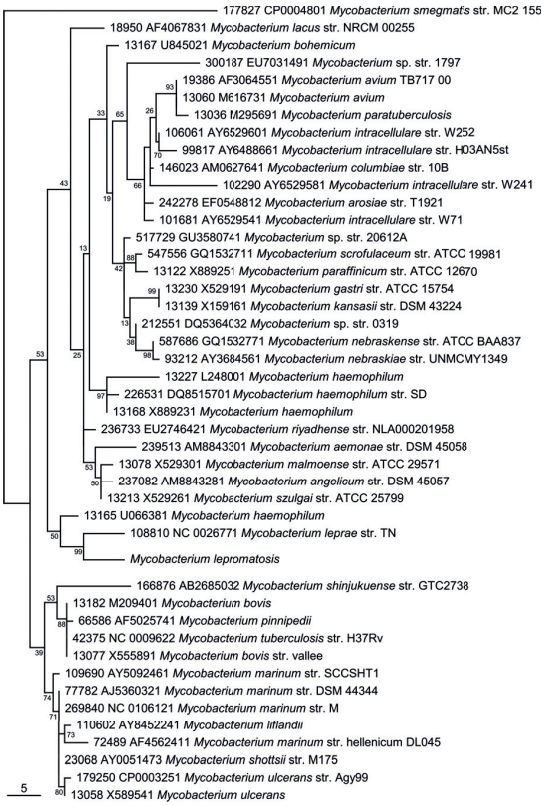


Figure S5: Phylogeny of selected mycobacterial species in a maximum parsimony tree based on 16S rRNA sequence alignments. Greengene number and accession number are given for each label. Bootstrap support is indicated for each node

Chapter 2.2 Insight from the genome sequence of *Mycobacterium lepraemurium:* massive gene decay and reductive evolution

Andrej Benjak,[a]* Tanvi P. Honap,[b]* Charlotte Avanzi,[a] Enrique Becerril-Villanueva,[c] Iris Estrada-García,[d] Oscar Rojas-Espinosa,[d] Anne C. Stone,[e,f #] Stewart T. Cole[a#]

[a]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Switzerland;
[b]School of Life Sciences, Arizona State University, Tempe, Arizona, USA;
[c]Departamento de Psicoinmunología, Instituto Nacional de Psiquiatría Ramón de la Fuente, Mexico City, Mexico
[d]Departamento de Inmunología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico City, Mexico
[e]School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA
[f]Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA

[*] these authors contributed equally to this work
[#] corresponding author

2017. mBio 8(5):e01283-17

Contributions: DNA extraction and library preparation

This is a pre-print, author-modified version of an article accepted for publication in mBio following peer review. The version of record is available online in free open access at:
doi: 10.1128/mBio.01283-17

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

Insights from the Genome Sequence of *Mycobacterium lepraemurium*: Massive Gene Decay and Reductive Evolution

Andrej Benjak, [a]* **Tanvi P. Honap** [b] * **Charlotte Avanzi,** [a] Enrique Becerril-Villanueva,[c] Iris Estrada-García,[d] Oscar Rojas-Espinosa,[d] Anne C. Stone,[e,f] Stewart T. Cole[a]

Global Health Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland[a]; School of Life Sciences, Arizona State University, Tempe, Arizona, USA[b]; Departamento de Psicoinmunología, Instituto Nacional de Psiquiatría Ramón de la Fuente, Mexico City, Mexico[c]; Departamento de Inmunología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Mexico City, Mexico[d]; School of Human Evolution and Social Change, Arizona State University, Tempe, Arizona, USA[e]; Center for Evolution and Medicine, Arizona State University, Tempe, Arizona, USA[f]

ABSTRACT *Mycobacterium lepraemurium* is the causative agent of murine leprosy, a chronic, granulomatous disease similar to human leprosy. Due to the similar clinical manifestations of human and murine leprosy and the difficulty of growing both bacilli axenically, *Mycobacterium leprae* and *M. lepraemurium* were once thought to be closely related, although it was later suggested that *M. lepraemurium* might be related to *Mycobacterium avium*. In this study, the complete genome of *M. lepraemurium* was sequenced using a combination of PacBio and Illumina sequencing. Phylogenomic analyses confirmed that *M. lepraemurium* is a distinct species within the *M. avium* complex (MAC). The *M. lepraemurium* genome is 4.05 Mb in length, which is considerably smaller than other MAC genomes, and it comprises 2,682 functional genes and 1,139 pseudogenes, which indicates that *M. lepraemurium* has undergone genome reduction. An error-prone repair homologue of the DNA polymerase III *α*-subunit was found to be nonfunctional in *M. lepraemurium*, which might contribute to pseudogene formation due to the accumulation of mutations in nonessential genes. *M. lepraemurium* has retained the functionality of several genes thought to influence virulence among members of the MAC.

IMPORTANCE *Mycobacterium lepraemurium* seems to be evolving toward a minimal set of genes required for an obligatory intracellular lifestyle within its host, a niche seldom adopted by most mycobacteria, as they are free-living. *M. lepraemurium* could be used as a model to elucidate functions of genes shared with other members of the MAC. Its reduced gene set can be exploited for studying the essentiality of genes in related pathogenic species, which might lead to discovery of common virulence factors or clarify host-pathogen interactions. *M. lepraemurium* can be cultivated *in vitro* only under specific conditions and even then with difficulty. Elucidating the metabolic (in)capabilities of *M. lepraemurium* will help develop suitable axenic media and facilitate genetic studies.

KEYWORDS *Mycobacterium lepraemurium*, comparative genomics, genome sequencing, murine leprosy

Murine leprosy is a chronic, granulomatous disease caused by *Mycobacterium lepraemurium*. Murine leprosy mainly affects the skin, mucosa of the upper respiratory tract, and eyes; however, unlike human leprosy, the viscera are commonly affected,

Received 22 July 2017 **Accepted** 11 September 2017 **Published** 17 October 2017

Citation Benjak A, Honap TP, Avanzi C, BecerrilVillanueva E, Estrada-García I, Rojas-Espinosa O, Stone AC, Cole ST. 2017. Insights from the genome sequence of *Mycobacterium lepraemurium*: massive gene decay and reductive evolution. mBio 8:e01283-17. https://doi.org/10.1128/mBio.01283-17.

Invited Editor Roland Brosch, Institut Pasteur

Editor Christina L. Stallings, Washington University in St. Louis School of Medicine

Address correspondence to Anne C. Stone, acstone@asu.edu, or Stewart T. Cole, stewart.cole@epfl.ch.

*Present address: Tanvi P. Honap, Department of Anthropology, University of Oklahoma, Norman, Oklahoma, USA.

A.B. and T.P.H. contributed equally to this work.

whereas the peripheral nerves are not (1). Murine leprosy was first reported in the early 20th century in rats in Ukraine (2), after which similar cases were reported from other countries (3, 4). *M. lepraemurium* is one of the causative agents of leprosy in cats and causes granulomatous skin lesions that often involve ulceration (5).

In humans, leprosy is primarily caused by *Mycobacterium leprae* and *Mycobacterium lepromatosis*. Numerous similarities exist between human and murine leprosy, including disease transmission through abrasions in the skin and the mucosal respiratory surfaces, a similar spectrum of disease progression, suppression of cell-mediated immunity, and strong humoral immunity (1).

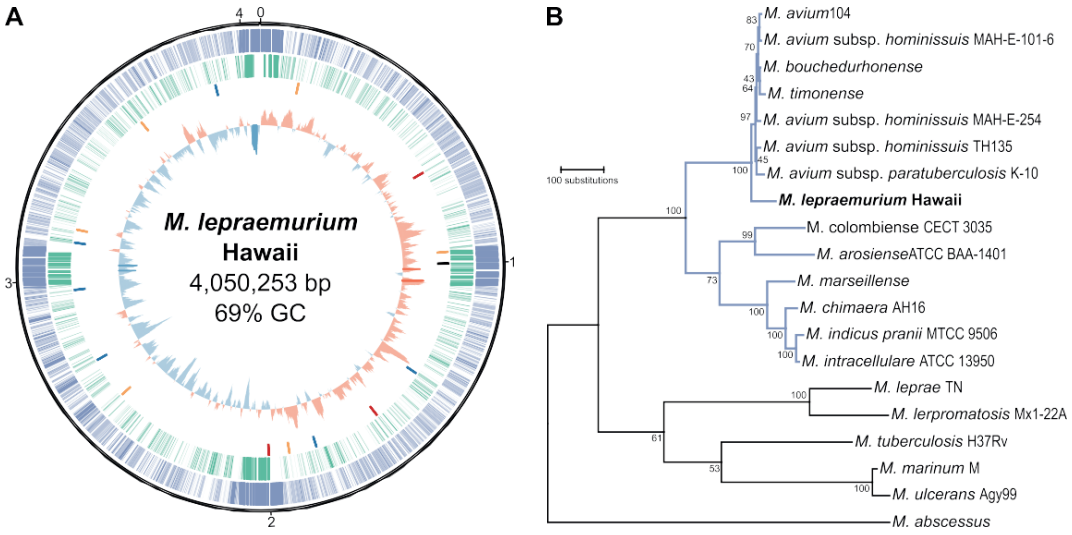


FIG 1 The genome of *Mycobacterium lepraemurium* strain Hawaii. (A) Graphical representation of the genome and its features. The origin of replication is at 12 o'clock, and the genome sequence runs clockwise. Ticks around the outermost circle mark million bases. Looking inwards, the outermost track or the first track (blue) shows functional genes. The second track (green) shows pseudogenes. The third track shows insertion sequences (all dysfunctional) colored to distinguish families (orange, red, black, and blue) and exaggerated in size for visibility. The fourth track shows the GC skew, calculated for a 20-kb window sliding every 1 kb, represented as a histogram with positive values pointing outward (red bars) and negative values pointing inward (blue bars). (B) Maximum parsimony tree of *M. lepraemurium* and selected mycobacterial species. The tree was created using MEGA7 from concatenated amino acid sequences (3,936 positions) of 11 proteins (DnaN, RplI, GrpE, MetG, RplY, PheT, FtsQ, HolA, MiaA, FtsY, and FtsX). Branches corresponding to the *Mycobacterium avium* complex are in blue. Bootstrap support, estimated from 500 replicates, is given below each branch. An expanded version of this tree, including additional genomes, is in Fig. S1 in the supplemental material.

This led to the hypothesis that these species were closely related, and hence, it was thought that murine leprosy might serve as a model for human leprosy (6–8). DNA hybridization studies and analysis of the 16S rRNA gene sequences suggested that *M. lepraemurium* might actually be more related to the *Mycobacterium avium* complex (MAC) (9, 10). However, to date, no phylogenomic study of *M. lepraemurium* has been conducted, and the lack of a genome sequence has restricted our understanding of its biology and evolution.

Here, we describe the complete genome of *M. lepraemurium*, which was sequenced using single-molecule real-time (SMRT; Pacific Biosciences) and Illumina technologies. The *M. lepraemurium* genome was found to be circular, with a total GC content of 68.99%, and 4,050,523 bp in length. No plasmids were found. The genome comprises 3,821 "protein-coding" genes, of which 2,682 are functional genes and 1,139 are pseudogenes (Fig. 1A). *M. lepraemurium* belongs to the MAC and is more closely related to the *M. avium* clade than to the *M. intracellulare* clade (Fig. 1B and see Fig. S1 and S2 in the supplemental material).

Genome downsizing and pseudogene formation. At 4.05 Mb, the *M. lepraemurium* genome is the smallest within the MAC and one of the smallest among mycobacteria (9th out of the 350 sequenced mycobacterial species). More strikingly, within the mycobacteria, only the *M. leprae* and *M. lepromatosis* genomes contain fewer functional protein-coding genes than the *M. lepraemurium* genome. The presence of 1,139 pseudogenes indicates that *M. lepraemurium* underwent reductive evolution, which is characteristic for strictly host-associated organisms. Analyses of pseudogene families within a diverse set of prokaryotes have shown that pseudogenes are most likely to occur in ABC transporter, short-chain dehydrogenase/reductase, sugar transporter, cytochrome P450, and proline-glutamate (PE)/proline-proline-glutamate (PPE) gene families (11). The *M. lepraemurium* genome was found to contain pseudogenes in all these families.

M. lepraemurium is the third mycobacterial species known to have undergone reductive evolution. The other species include the common ancestor of *M. leprae* and the closely related *M. lepromatosis*, which underwent genome reduction before the two species diverged, as well as *Mycobacterium ulcerans*, which is in a state of intermediate reductive genome evolution (12–14). Genome size among *M. avium* subspecies varies considerably (around 4.8 to 5.5 Mb), but no extensive pseudogenization was observed in these organisms. Curiously, *M. lepraemurium* seems to be evolving convergently toward a minimal gene set such as the one retained by *M. leprae* (Fig. 2). This may be due to both species having adapted to a similar niche, consequently resulting in similar pathological manifestations.

Loss of the DnaQ-mediated proofreading mechanism of DNA polymerase III *a*-subunit has been hypothesized as the cause of pseudogene formation in *M. leprae* (12), and a similar pattern has been observed in other obligate pathogens and symbionts (15, 16). In *M. lepraemurium*, a homologue of the error-prone repair DNA polymerase III *a*-subunit (*MLM_3495*) is nonfunctional, which might contribute to a higher error rate leading to pseudogene formation in this species.

The AT content of the pathogen/symbiont genome seems to correlate with the age of the strict association with the host (16). As expected, the genome of *M. leprae* has the highest AT content of all mycobacteria (42.2%), especially in pseudogenes (43.5%). On the other hand, the genome of *M. lepraemurium* is as AT-poor as that of its closest relative, *M. avium* (31%), and shows only a small difference in the AT content of functional genes and pseudogenes (30.46% and 31.05%, respectively). This suggests that *M. lepraemurium* adopted the "strictly intracellular" niche relatively recently, as evidenced by its phylogenetic relatedness to the free-living *M. avium* (Fig. S1). Genomic data for other *M. lepraemurium* strains would allow us to determine whether the pseudogene content varies between strains, which would indicate ongoing reductive evolution.

Repetitive sequences and mobile elements. Repeats and mobile elements are the major vehicles leading to genomic rearrangements and large deletions, and hence, they are thought to play an important role in genome downsizing. Yet, there seems to be a negative correlation between the stage of reductive evolution and repeat content (17). This is likely because deletions and pseudogenization impact fitness as more genes become nonfunctional, and the remaining genes become indispensable. Therefore, it has been suggested that repeats might be involved in the early phase of reductive evolution and that a different mechanism is responsible for the gradual deletion of pseudogenes (17). *M. lepraemurium* has only a few repeats (Table S1), which mostly derive from three families of dysfunctional insertion sequences (IS), each consisting of up to six copies.

Curiously, one IS family consisting of six copies (*MLM_1414*, *MLM_1873*, *MLM_2691*, *MLM_2913*, *MLM_3078*, and *MLM_3876*) shows 84% nucleotide identity with the ISMsm2 mobile element from *Mycobacterium smegmatis* (NCBI accession number WP_ 003887303). A BLAST search showed that this IS family resides in only a few other unrelated mycobacteria and not in the MAC, with the exception of a more diverged copy (with 78% identity) present in the plasmid of *M. avium* subsp. *hominissuis* strain 88Br (GenBank accession number KR997898.1) and a truncated copy (with 81% identity) in the genome of *M. avium* subsp. *hominissuis* strain H87 (GenBank accession number CP018363.1).

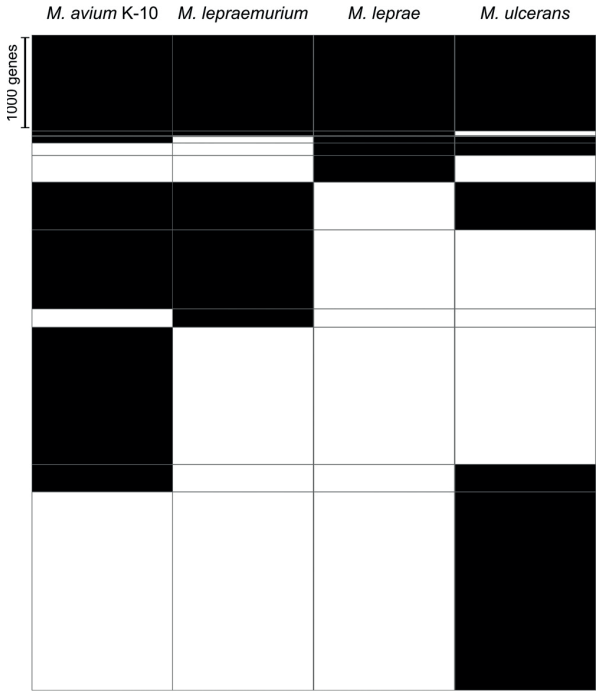


FIG 2 Heatmap of the gene orthology between *M. avium* subsp. *paratuberculosis* K-10, *M. lepraemurium*, *M. leprae*, and *M. ulcerans*. Genes are shown in black, and the absence of genes is shown in white. The raw height corresponds to the number of genes.

Strikingly, when we analyzed the genomic synteny between *M. lepraemurium*, *M. intracellulare*, and *M. avium* subsp. *hominissuis*, all the synteny breaks in *M. lepraemurium* occurred at the locations of this particular IS (Fig. S3A and B). The synteny plot between *M. lepraemurium* and *M. avium* subsp. *paratuberculosis* looked different from the aforementioned genomes, which is probably due to genomic rearrangements in *M. avium* subsp. *paratuberculosis* (Fig. S3C).

The presence of identical copies of dysfunctional transposable elements indicates that *M. lepraemurium* lost functional transposases relatively recently. As mentioned above, at least one IS family was responsible for large genomic rearrangements in *M. lepraemurium*, and a deeper analysis of the genome sequence might shed light on the driving force that caused major deletions in this genome.

***M. lepraemurium*-specific sequences.** Comparison of *M. lepraemurium* with other members of the MAC revealed a total of 32.2 kb of genomic sequence that was not found in any publicly available genome sequence. These regions contain 35 genes, of which only 12 are functional (Table S2). One of these genes, *MLM_3300*, codes for a Fic family protein. Fic proteins have been associated with pathogenicity in bacteria, often acting as toxins that interfere with the host cell in different ways (18).

Interaction with macrophages. After entering the macrophage, members of the MAC reside within phagosomes (19). There they inhibit phagosome maturation by preventing acidification to a pH below 6.4, and thus, prevent fusion with the extremely acidic lysosome. In *M. avium*, mutations in the PPE gene *MAV_2928* and the PE gene *MAV_1346* inhibit maturation and acidification of phagosomes, resulting in decreased virulence (20, 21). In *M. lepraemurium*, gene *MLM_2357* (homologous to *MAV_2928*) and *MLM_1265* (homologous to *MAV_1346*) are functional, suggesting that they may help its survival in macrophages. Additionally, the gene *MLM_2012* encodes the lipoprotein, LppM, which is an important virulence factor in *M. tuberculosis* and interferes with phagosomal maturation in macrophages (22).

Reactive oxygen species (ROS) are produced by the macrophage as a bactericidal

mechanism in response to infection. The ability of pathogens to produce enzymes such as catalase-peroxidase, epoxide hydrolase, and superoxide dismutase (SOD), which remove ROS, enable their survival within macrophages. While it has been suggested that *M. lepraemurium* abolishes the production of ROS upon entering the macrophage (29), probably as a survival mechanism, it remains unclear whether small quantities of ROS are produced upon infection and handled by the bacterium. *M. lepraemurium* has retained four functional epoxide hydrolases (*MLM_0642*, *MLM_0684*, *MLM_1194*, and *MLM_1485*) and one catalase-peroxidase (*MLM_2092*), which explains its observed catalase-peroxidase activity (23). Moreover, *M. lepraemurium* is able to produce two superoxide dismutases, SodA and SodC (*MLM_0123* and *MLM_2650*).

Glycopeptidolipid synthesis. Glycopeptidolipids (GPLs) are synthesized by several nontuberculosis mycobacteria, including members of the MAC. GPLs are present on the outermost layer of the cell wall, and therefore, they play an important role in sliding motility, biofilm formation, and pathogenesis (24). This is even more relevant within the MAC, whose members can produce a number of serovar-specific variations of GPLs. The *M. lepraemurium* cell wall contains glycolipids and amino acids (25, 26); however, production of GPLs in *M. lepraemurium* has not been studied. Most of the genes known to be involved in the biosynthesis of non-serovar-specific GPLs are intact in *M. lepraemurium*, albeit with a few exceptions (Table S3). Although it is difficult at this moment to predict the effects of these mutations, it appears that *M. lepraemurium* should be able to produce GPLs, as all of the core genes are present.

Virulence. While the ESX-1 system is the main determinant of virulence in *M. tuberculosis* and *M. leprae* (27), it is absent in the MAC. However, ESX-5, which is also associated with virulence in pathogenic mycobacteria, is mostly intact in *M. lepraemurium*, except for the cytochrome P450 hydroxylase (*MLM_2361*), which is also nonfunctional in *M. leprae*. The duplicated four-gene region, ESX-5a, encoding EsxI, EsxJ, a PPE, and a PE protein may serve as an accessory system for transport of a subset of ESX-5 proteins (28) and is still functional in *M. lepraemurium*. In *M. lepraemurium*, the PPE and PE genes are merged into a single open reading frame, as in *M. avium* subsp. *paratuberculosis* but not in other members of the MAC. An interesting observation is that in MAC, *esxI* and *esxJ* are 100% identical to *esxN* and *esxM* from the main ESX-5 locus (unlike in *M. tuberculosis*), which indicates a novel and recent duplication event and suggests that a crucial function might lie behind redundancy of the ESX-5 components.

The *pks12* gene is involved in the biosynthesis of mannosyl-*13*-1-phosphomycoketides (MPM) and is found only in the slow-growing mycobacteria. In different pathogenic mycobacteria, including *M. avium* (20), *pks12* was shown to be necessary for the virulence, and this gene is functional in *M. lepraemurium* (*MLM_2156*).

Experimental procedures. *M. lepraemurium* strain Hawaii was grown in BALB/c mice. The DNA was sequenced using Illumina and PacBio technologies, followed by sequence assembly and annotation. More details are given in Text S1 in the supplemental material.

Accession number(s). The annotated genome was submitted to GenBank under GenBank accession number CP021238.

ACKNOWLEDGMENTS

We thank the Genomic Technologies Facility at the University of Lausanne, Switzerland, for Illumina and PacBio sequencing. We thank Philippe Busso (École Polytech nique Fédérale de Lausanne [EPFL], Lausanne, Switzerland) for helping with the DNA extraction. This work was supported by the Fondation Raoul Follereau and the Swiss National Science Foundation grant IZRJZ3_164174.

REFERENCES

1. Rojas-Espinosa O. 2009. Murine leprosy revisited, p 97–140. In Tomioka H (ed), Current topics on the profiles of host immunological response to mycobacterial infections, 1st ed. Research Signpost, Trivandrum, Kerala, India.
2. Stefansky WK. 1902. Zabolevanija u krys, vyzvannyja kislotoupornoj palotsjkoj. Rus Vratsj 47:1726 –1727.
3. Marchoux E, Sorel F. 1912. Recherches sur la lèpre: la lèpre des rats (Lepra murium). Ann Inst Pasteur 56:778 – 801.
4. Dean G. 1903. A disease of the rat caused by an acid-fast bacillus. Cent F Bakteriol

34:222–224.
5. O'Brien CR, Malik R, Globan M, Reppas G, McCowan C, Fyfe JA. 2017. Feline leprosy due to *Mycobacterium lepra*emurium. J Feline Med Surg 19:737–746. https://doi.org/10.1177/1098612X17706469.
6. Walker EL, Sweeney MA. 1929. The identity of human leprosy and rat leprosy. J Prev Med 3:325–333.
7. Schmitt LS. 1911. On the relation between rat and human leprosy. Pathologe 2:29 –37.
8. Dean G. 1905. Further observations on a leprosy-like disease of the rat. J Hyg 5:99 – 112. https://doi.org/10.1017/S0022172400002370.
9. Athwal RS, Deo SS, Imaeda T. 1984. Deoxyribonucleic acid relatedness among *Mycobacterium leprae*, *Mycobacterium lepra*emurium, and selected bacteria by dot blot and spectrophotometric deoxyribonucleic acid hybridization assays. Int J Syst Bacteriol 34:371–375. https://doi.org/ 10.1099/00207713-34-4-371.
10. Hughes MS, Ball NW, Beck LA, De Lisle GW, Skuce RA, Neill SD. 1997. Determination of the etiology of presumptive feline leprosy by 16S rRNA gene analysis. J Clin Microbiol 35:2464 –2471.
11. Liu Y, Harrison PM, Kunin V, Gerstein M. 2004. Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. Genome Biol 5:R64. https://doi .org/10.1186/gb-2004-5-9-r64.
12. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D, Mungall K, Basham D, Brown D, Chillingworth T, Connor R, Davies RM, Devlin K, Duthoy S, Feltwell T, Fraser A, Hamlin N, Holroyd S, Hornsby T, Jagels K, Lacroix C, Maclean J, Moule S, Murphy L, Oliver K, Quail MA, Rajandream M-A, Rutherford KM, Rutter S, Seeger K, Simon S, Simmonds M, Skelton J, Squares R, Squares S, Stevens K. 2001. Massive gene decay in the leprosy bacillus. Nature 409:1007–1011. https://doi.org/10.1038/35059006.
13. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, Nieselt K, Krause J, Vera-Cabrera L, Cole ST. 2015. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proc Natl Acad Sci U S A 112:4459 – 4464. https://doi.org/ 10.1073/pnas.1421504112.
14. Stinear TP, Seemann T, Pidot S, Frigui W, Reysset G, Garnier T, Meurice G, Simon D, Bouchier C, Ma L, Tichit M, Porter JL, Ryan J, Johnson PDR, Davies JK, Jenkin GA, Small PLC, Jones LM, Tekaia F, Laval F, Daffé M, Parkhill J, Cole ST. 2007. Reductive evolution and niche adaptation inferred from the genome of Mycobacterium ulcerans, the causative agent of Buruli ulcer. Genome Res 17:192–200. https://doi.org/10.1101/ gr.5942807.
15. Dale C, Wang B, Moran N, Ochman H. 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. Mol Biol Evol 20:1188 –1194. https://doi.org/10.1093/molbev/msg138.
16. Moya A, Peretó J, Gil R, Latorre A. 2008. Learning how to live together: genomic insights into prokaryote-animal symbioses. Nat Rev Genet 9:218 –229. https://doi.org/10.1038/nrg2319.
17. Frank AC, Amiri H, Andersson SGE. 2002. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. Genetica 115:1–12. https://doi.org/10.1023/A:1016064511533.
18. Roy CR, Cherfils J. 2015. Structure and function of Fic proteins. Nat Rev Microbiol 13:631– 640. https://doi.org/10.1038/nrmicro3520.
19. Ignatov D, Kondratieva E, Azhikina T, Apt A. 2012. Mycobacterium avium-triggered diseases: pathogenomics. Cell Microbiol 14:808 – 818. https://doi.org/10.1111/j.1462-5822.2012.01776.x.
20. Li Y-J, Danelishvili L, Wagner D, Petrofsky M, Bermudez LE. 2010. Identification of virulence determinants of Mycobacterium avium that impact on the ability to resist host killing mechanisms. J Med Microbiol 59:8 –16. https://doi.org/10.1099/jmm.0.012864-0.
21. Li Y, Miltner E, Wu M, Petrofsky M, Bermudez LE. 2005. A Mycobacterium avium PPE gene is associated with the ability of the bacterium to grow in macrophages and virulence in mice. Cell Microbiol 7:539 –548. https:// doi.org/10.1111/j.1462-5822.2004.00484.x.
22. Deboosère N, Iantomasi R, Queval CJ, Song OR, Deloison G, Jouny S, Debrie AS, Chamaillard M, Nigou J, Cohen-Gonsaud M, Locht C, Brodin P, Veyron-Churlet R. 2017. LppM impact on the colonization of macrophages by *Mycobacterium tuberculosis*. Cell Microbiol 19:e12619. https://doi.org/10.1111/cmi.12619.
23. Lygren ST, Closs O, Bercouvier H, Wayne LG. 1986. Catalases, peroxidases, and superoxide dismutases in *Mycobacterium leprae* and other mycobacteria studied by crossed immunoelectrophoresis and polyacrylamide gel electrophoresis. Infect Immun

54:666 – 672.
24. Pang L, Tian X, Pan W, Xie J. 2013. Structure and function of mycobacterium glycopeptidolipids from comparative genomics perspective. J Cell Biochem 114:1705–1713. https://doi.org/10.1002/jcb.24515.
25. Draper P. 1971. The walls of *Mycobacterium lepra*emurium: chemistry and ultrastructure. J Gen Microbiol 80:313–324.
26. Luna-Herrera J, Rojas-Espinosa O, Estrada-Parra S. 1996. Recognition of lipid antigens by sera of mice infected with *Mycobacterium lepra*emurium. Int J Lepr Other Mycobact Dis 64:299 –305.
27. Gröschel MI, Sayes F, Simeone R, Majlessi L, Brosch R. 2016. ESX secretion systems: mycobacterial evolution to counter host immunity. Nat Rev Microbiol 14:677– 691. https://doi.org/10.1038/nrmicro.2016.131.
28. Shah S, Briken V. 2016. Modular organization of the ESX-5 secretion system in *Mycobacterium tuberculosis*. Front Cell Infect Microbiol 6:49. https://doi.org/10.3389/fcimb.2016.00049.
29. Rojas-Espinosa O, Camarena-Servin V, Estrada-Garcia I, Arce-Paredes P, Wek-Rodriguez K. 1998. *Mycobacterium lepra*emurium, a well-adapted parasite of macrophages: I. Oxygen metabolites. Int J Lepr Other Mycobact Dis 66:365–

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio

Bacilli culture and purification. *M. lepraemurium* Hawaii was grown using serial infections in BALB/c mice injected by the intraperitoneal route. At four to six months post-infection, the infected spleen and liver were harvested. Bacteria were purified by following the protocol in (1), followed by the Percoll step (2) and then by using previously established protocols (3, 4). Briefly, 4 g of tissue was suspended in 20 ml of 0.2 M sucrose and ground in a glass Potter–Elvehjem homogenizer. The resulting suspension was centrifuged for 20 min at 150 xg to separate cell debris (Sorvall RC5B, rotor HB4, Sorvall Instruments, Wilmington, Delaware, USA). Then, 9 ml of the isolated supernatant was overlaid onto 12 ml of 0.3 M sucrose and the tubes were centrifuged at 3,500 xg for 10 min at 4–10°C (Sorvall RC5B, rotor SS34). The resulting bacilli-rich pellet was resuspended in 20 ml of 0.2 M sucrose and overlaid, in 9-ml aliquots, onto 12 ml of 1.5 M KCl. The tube was then centrifuged at 4°C for 10 min at 3,500 xg. The bacilli were collected, washed 3 times with phosphate-buffered-saline (PBS) at pH 7.4 (PBS is 0.01 M Na/K phosphate, 0.15 M NaCl), and suspended in 40 ml of a solution containing a mixture of Percoll (3 parts) and 0.1% Tween 80 (7 parts). The suspension was centrifuged at 23,000 xg (Sorvall RC5B, rotor SS34) for 60 min at 40°C. Then the bacillary layer was resuspended in 20 ml of Percoll-Tween and centrifuged as before. The final bacillary pellet was collected and washed 5 times with PBS pH 7.4 or until the Percoll was completely eliminated. The purity of the bacillary preparation was verified by Ziehl–Neelsen staining. The purified bacillary suspension was prepared in synthetic 7H9 Middlebrook broth-OADC medium (DIFCO, Detroit, MI, USA), and quantified via a nephelometric reference curve prepared with known quantities of bacteria. The bacillary suspension was aliquoted and frozen at -20°C until ready for use.
For DNA extraction, several aliquots were combined, centrifuged at 3,500 xg for 10 min, the supernatant was eliminated and the pellet was frozen (-20°C) without any further treatment.

DNA extraction. DNA extraction was carried out using a custom-designed protocol for mycobacterial DNA. The bacterial cell pellet was washed with 500 µL of phosphate buffer saline (PBS) prior to centrifugation at 5000 g for 10 min. The supernatant was discarded and the pellet was re-suspended in 1 mL of bacterial lysis buffer B1 (50 mM Tris-HCl pH 8.0; 50 mM EDTA pH 8.0; 0.5% Tween 20; 0.5% Triton-X100) containing 45 µL of proteinase K (20 mg/mL) and 20 µL of lysozyme (100 mg/mL). The mixture was then transferred into bead-beating tubes containing 500 µL of silica beads (0.1 mm zirconia beads) prior to physical disruption using the Precellys24 homogenizer at 6.5 m/s for 25 sec. After incubating at 56°C for 1 h, the mixture was centrifuged and the supernatant was transferred to a new tube. An additional incubation with 20 µL proteinase K (20 mg/mL) was conducted at 56°C for 30 min. The mixture was then incubated at 4°C for 15 min. RNAse A (Sigma) was added and the sample was incubated 30 min at 37°C, followed by the addition of 350 µL of bacterial lysis buffer B2 (3M guanidine hydrochloride, 20% Tween 20), and incubated for 30 min at 50°C. DNA was purified using the Qiagen Genomic-Tip/20G according to the manufacturer's instructions, and eluted in 2 mL elution buffer. The DNA was precipitated using 0.7x volume of isopropanol and centrifuged at 4°C for 15 min. The pellet was washed twice with 200 µL 70% ethanol, air-dried, and suspended

overnight in 200 µL Tris HCl buffer (pH 8.0) at room temperature under continuous shaking. The DNA was then purified using AMPure beads (Thermofisher) at a ratio of 0.45. The quality of the DNA extract was checked using the Fragment Analyzer (Advanced Analytical Technologies) and quantified using the Qubit 2.0 (Life Technologies).

Illumina sequencing. DNA (50 µL) was sheared using the Covaris S220 Focused-ultrasonicator (Covaris) to obtain 400 bp-long DNA fragments, and purified using AMPure beads (1.8x) and the manufacturer's protocol. The sheared DNA was quantified using the dsDNA High Sensitivity assay and the Qubit 2.0 flurometer (Life Technologies). Up to 1 µg of DNA in 50 µL was used for library preparation using the Kapa Hyper prep kit (Roche) and PentAdapters (Pentabase) for indexing. The library was quantified using the dsDNA Broad Range assay and the Qubit 2.0 fluorometer. The library was sequenced on an Illumina HiSeq 2500 (1 x 101 bp run).

PacBio sequencing. DNA (5.1 µg) was sheared using a Covaris g-TUBE (Covaris S220) to obtain 10 kb fragments and the size distribution was checked using the Fragment Analyzer (Advanced Analytical Technologies). Sheared DNA (4 µg) was used to prepare a SMRTbell library with the PacBio SMRTbell Template Prep Kit 1 (Pacific Biosciences) according to the manufacturer's recommendations. The resulting library was size-selected using a BluePippin system (Sage Science, Inc.) for molecules larger than 8 kb. The recovered library was sequenced using a SMRT cell with P6/C4 chemistry and MagBeads on a PacBio RSII system (Pacific Biosciences) at 240 min movie length.

Genome assembly. PacBio reads were processed using the HGAP2 and HGAP3 pipelines (5). The resulting contigs were compared to the nucleotide database at NCBI using BLAST (6). The two largest contigs produced by HGAP3 v2.3.0 (which were 2.3 and 1.7 Mb in length, respectively) matched to *M. avium* sequences. These two contigs corresponded to the three largest contigs produced by HGAP2 v2.3.0 (which were 1.7, 1.6, and 0.6 Mb in length) and two shorter contigs (61 and 21 kb in length). The two HGAP3 contigs could be joined by the overlapping HGAP2 contigs, resulting in a single consensus sequence with overlapping ends, indicative of a circular genome. To correct for possible sequence errors, Illumina reads were mapped onto the draft genome sequence using Bowtie2 (7) resulting in 35-fold coverage of non-duplicate reads. Variants were called using SAMtools mpileup (Li et al., 2009) and VarScan2 (9), resulting in only five single nucleotide polymorphisms (SNPs) and two short insertion-deletions (InDels).
Illumina reads (4%) that did not map to the final genome sequence were assembled using MIRA (https://sourceforge.net/projects/mira-assembler/). The resulting 34 contigs (of which the largest was 1.9 kbp long) were compared to the nucleotide and protein databases at NCBI using BLAST. The contigs matched to *Mus musculus* or to various bacteria. No evidence of a putative plasmid sequence was found.

Gene prediction. *De novo* gene prediction was conducted using the RAST server (10) with the frameshift correction option. Reference-based gene prediction was conducted using RATT (11) with annotations from *M. avium* subsp. *paratuberculosis* K-10 (NC_002944.2) and *M. avium* subsp. *hominissuis* TH135 (AP012555.1). All predictions were merged, and inconsistencies and large intergenic areas were manually checked by using BLAST to compare the problematic sequences against the protein database at NCBI. Gene predictions, shorter than 100 nucleotides in length and not conserved in the genomes of other *M. avium* species, were removed. The annotated genome was submitted to GenBank (accession number CP021238).

Phylogenetic analyses. Two different strategies were used for the phylogenetic analysis to assure accuracy and consistency of the reconstructed tree topologies, as described below.

Phylogenetic analysis of concatenated amino-acid sequences. A set of eleven genes was chosen for the analysis (12), (DnaN, RplI, GrpE, MetG, RplY, PheT, FtsQ, HolA, MiaA, FtsY, FtsX). To make sure to identify strains that are closely related to *M. lepraemurium*, the nucleotide sequence of *M. lepraemurium* corresponding to the abovementioned genes was used as dcBLASTn query against all available genome sequences and genome assemblies of the MAC available at NCBI as of June 2017. BLAST hits were translated into amino-acid sequence. Amino-acid sequences of additional genomes from the MAC complex as well as other mycobacteria were included. Analysis of the concatenated alignments was done in MEGA7 (13).

Phylogenetic analysis of whole-genome alignments. Publically available genome data were acquired for comparative purposes for 16 mycobacterial species. Contigs or finished genomes of these species were aligned to the *M. avium* 104 reference genome using LAST

(14) with the following parameters: -u = 0, -e = 34, and -j = 5. The maf-convert program was used to covert the alignment file to a SAM file and SAMtools was used obtain a BAM file which was used for further analyses. SAMtools mpileup and bcftools call were used to produce the VCF files. VCF files for all strains were combined using the CombineVariants tool available in the Genome Analysis Toolkit (GATK) (15). The SelectVariants tool in GATK was used to output a VCF file containing the sites comprising SNPs. VCFtools (16) was used to remove InDels, tri-allelic sites, and sites with missing data. An SNP alignment was generated using a publically available perl script (17), which comprised a total of 460,625 sites.

Phylogenetic trees were constructed using the Maximum Likelihood (ML) method in RAxML v7.2.8 (18) and the Neighbor-Joining (NJ) and Maximum Parsimony (MP) methods in MEGA7 (13). The ML tree was generated using the GTR-GAMMA model and 100 bootstrap replicates (Figure 1). The NJ tree was generated using the p-distance method and bootstrap support was estimated from 500 replicates (Figure S1). The MP tree was generated using the Subtree-Pruning-Regrafting (SPR) algorithm and 500 bootstrap replicates (Figure S2).

Comparison of orthologous genes. Orthologs between *M. lepraemurium* and *M. avium* subsp. *paratuberculosis* were inferred by RATT and manually during the annotation process. Orthologs between *M. avium*, *M. leprae* and *M. ulcerans* were retrieved from http://www.pathogenomics.sfu.ca/ortholugedb/

References:

1. Prabhakaran K, Harris EB, Kirchheimer WF. 1976. Binding of 14C labeled dopa by *Mycobacterium leprae* in vitro. Int J Lepr 44:58–64.
2. Draper P. 1980. Purification of *Mycobacterium leprae*Report of the Fifth Meeting of the Scientific Working Group on the Immunology of Leprosy, TDR/IMMLEP-SWG 5/80.3.
3. Wek-Rodriguez K, Silva-Miranda M, Arce-Paredes P, Rojas-Espinosa O. 2007. Effect of reactive oxygen intermediaries on the viability and infectivity of *Mycobacterium leprae*murium. Int J Exp Pathol 88:137–145.
4. Rojas-Espinosa O, Wek-Rodriguez K, Arce-Paredes P. 2002. The effect of exogenous peroxidase on the evolution of murine leprosy. IntJLeprOther MycobactDis 70:191–200.
5. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569.
6. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410.
7. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359.
8. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.
9. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22:568–576.
10. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75.
11. Otto TD, Dillon GP, Degrave WS, Berriman M. 2011. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res 39:e57.
12. Mizuno T, Natori T, Kanazawa I, Eldesouky I, Fukunaga H, Ezaki T. 2016. Core housekeeping proteins useful for identification and classification of mycobacteria 32.
13. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.
14. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. Genome Res 21:487–493.
15. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–303.

16. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.
17. Bergey C. 2012. vcf-tab-to-fasta.
18. Stamatakis A. 2006. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22:2688–2690.

Figure S1: Maximum parsimony tree of *M. lepraemurium* and other mycobacterial species. The tree was created using MEGA7 from concatenated amino acid sequences (3,948 positions) of 11 proteins (DnaN, RplI, GrpE, MetG, RplY, PheT, FtsQ, HolA, MiaA, FtsY, and FtsX). Bootstrap support, estimated from 500 replicates, is given below each branch.

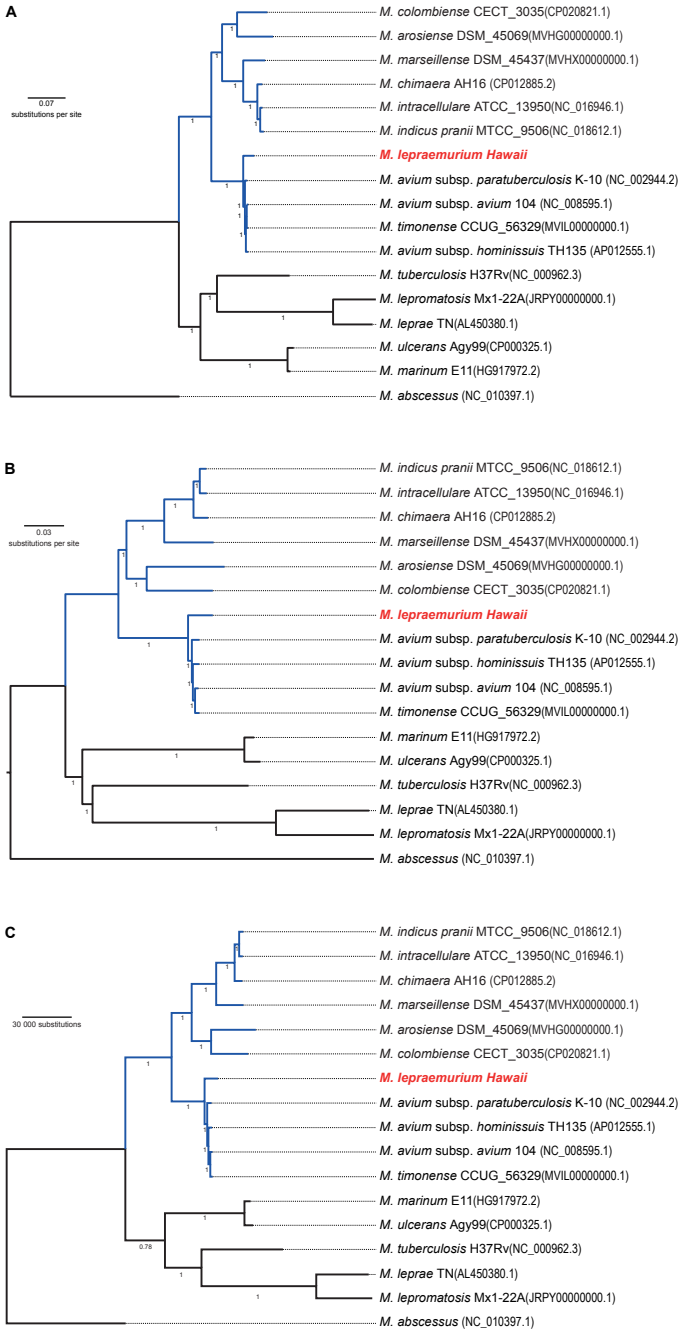


Figure S2: Phylogeny of *M. lepraemurium* and selected mycobacterial species based on whole-genome sequence alignments. Species belonging to the M. avium complex are highlighted in blue and *M. lepraemurium* is denoted in red. *M. abscessus* was used as the outgroup. (A) Maximum likelihood tree. The tree was created using RAxML based on 460,625 variable nucleotide sites and a general time reversible (GTR) model with gamma distribution. Bootstrap support estimated from 100 replicates is given below each branch. (B) Neighbor-joining tree. The tree was created using MEGA7 based on 460,625 variable nucleotide sites and the p-distance method. Bootstrap support estimated from 500 replicates is given below each branch. (C) Maximum parsimony tree. The tree was created using MEGA7 based on 460,625 variable nucleotide sites and the SPR algorithm. Bootstrap support estimated from 500 replicates is given below each branch.

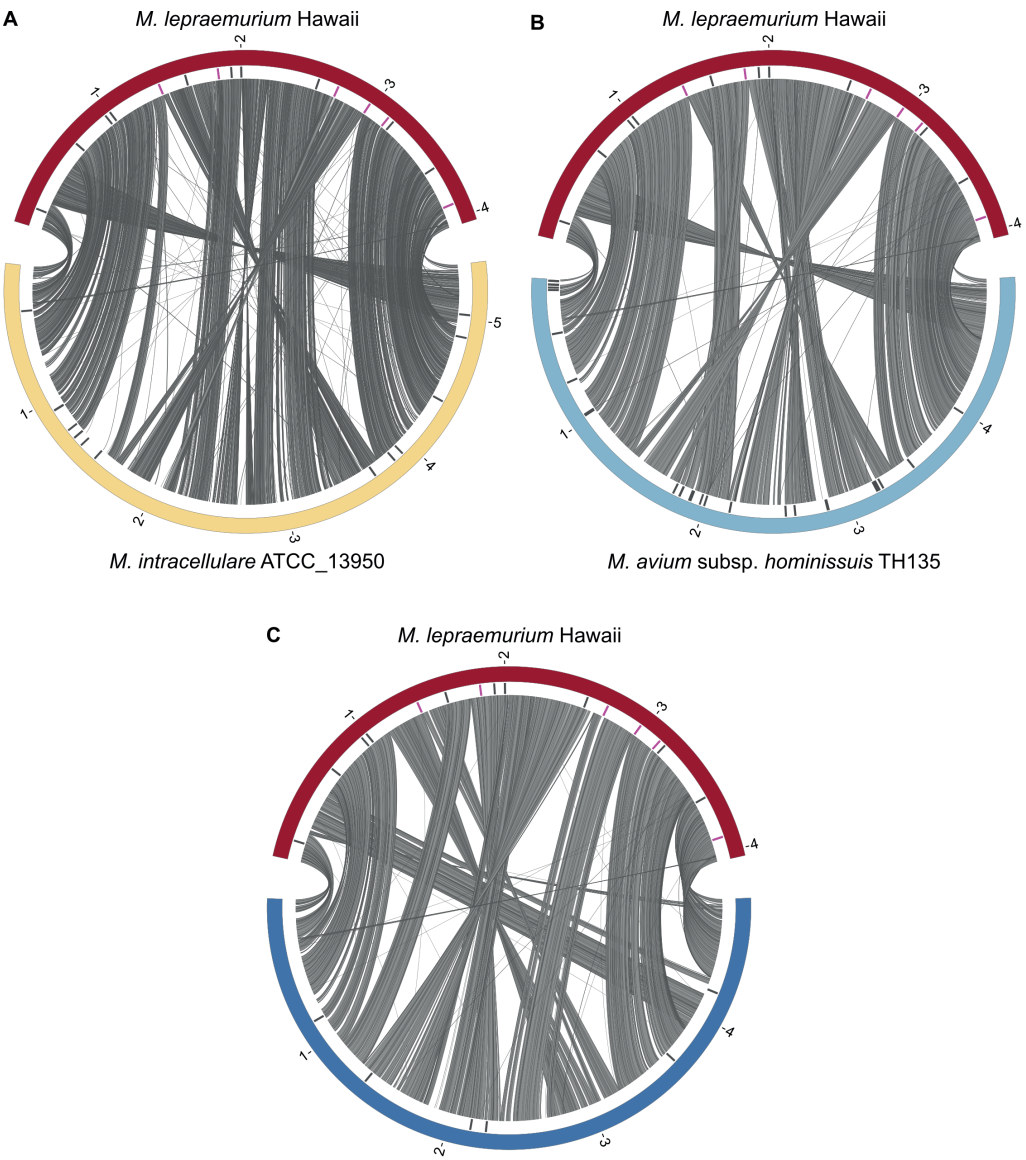


Figure S3: Synteny plots between *M. lepraemurium* and selected members of the Mycobacterium avium complex. Genome sequences are represented as colored ideograms. Numbered ticks mark million bases. Transposable elements are shown as ticks in the second track, with the ISMsm2-like family in magenta. Inner links are LAST hits (see Text S1 in the supplemental material for details).

Table S1: *M. lepraemurium* repetitive genomic regions.

Repeat cluster	Start	End	Length	Gene(s)	Pseudogene	Annotation
1	2590876	2593150	2275	MLM_2604	no	non-ribosomal peptide synthetase 1
	2595255	2597610	2356	MLM_2605	no	non-ribosomal peptide synthetase 2
	2599761	2602116	2356	MLM_2605	no	non-ribosomal peptide synthetase 2
2	2104068	2105674	1607	MLM_2156	no	malonyl CoA-acyl carrier protein transacylase
	2110167	2111773	1607	MLM_2156	no	malonyl CoA-acyl carrier protein transacylase
3	1337854	1339349	1496	MLM_1358	yes	partial REP13E12 repeat protein
	3383223	3384718	1496	MLM_3341	yes	uncharacterized protein (HNH endonuclease, 13E12 repeat family)
	3520566	3522018	1453	MLM_3478	yes	REP13E12 repeat protein (fragment)
4	672122	673674	1553	MLM_0680	yes	mobile element protein
	1596190	1597583	1394	MLM_1641	yes	IS481 family transposase
	1998427	1999979	1553	MLM_2038	yes	mobile element protein
5	2000175	2001365	1191	MLM_2040	yes	partial REP13E12 repeat protein
	2200411	2201601	1191	MLM_2245	yes	partial REP13E12 repeat protein
6	1398771	1399903	1133	MLM_1414	yes	mobile element protein
	1832525	1833069	545	MLM_1873	yes	mobile element protein
	2707760	2708895	1136	MLM_2691	yes	mobile element protein
	2962176	2963335	1160	MLM_2913	yes	mobile element protein
	3125465	3125919	455	MLM_3078	yes	mobile element protein
	3896248	3897425	1178	MLM_3876	yes	mobile element protein
7	1876090	1877213	1124	MLM_1920	no	3-oxoacyl-[acyl-carrier-protein] synthase
	3545427	3546550	1124	MLM_3503	no	3-oxoacyl-[acyl-carrier-protein] synthase
8	129992	131015	1024	MLM_0130	yes	mobile element protein
	959494	960036	543	MLM_0985	yes	mobile element protein
	1927549	1928568	1020	MLM_1967A	yes	mobile element protein
	2561454	2562473	1020	MLM_2584	yes	mobile element protein
	3170319	3171327	1009	MLM_3126	yes	mobile element protein
	3609141	3610160	1020	MLM_3568	yes	mobile element protein
9	1101447	1102085	639	MLM_1123 + MLM_1124	no	EsxI + EsxJ
	2311416	2312054	639	MLM_2353 + MLM_2354	no	EsxN + EsxM
10	3382732	3383042	311	MLM_3339	no	hypothetical protein
	3383043	3383353	311	MLM_3341	yes	uncharacterized protein (HNH endonuclease, 13E12 repeat family)

Table S2: ***M. lepraemurium*** **specific genomic regions and genes.**

Genomic region	Length	Overlapping gene	Start	Stop	Product	Function
954701- 956100	1400	MLM_0980	954318	955445	FAD-binding monooxygenase	Protein-coding
		MLM_0981	955549	956094	TetR family transcriptional regulator	Protein-coding
1039201- 1039800	600	NA	NA	NA	NA	NA
1042001- 1042600	600	MLM_1065	1042056	1042322	Helix-turn-helix XRE-family transcriptional regulator	Pseudogene
		MLM_1066	1042397	1042630	hypothetical protein	Protein-coding
1308101- 1308400	300	MLM_1327	1308059	1308490	putative extradiol dioxygenase	Protein-coding
1443701- 1444000	300	MLM_1462	1443654	1444794	uncharacterized protein	Pseudogene
1789901- 1792100	2200	MLM_1829	1789973	1790841	restriction endonuclease	Pseudogene
		MLM_1829A	1791008	1791931	hypothetical protein	Protein-coding
2022601- 2024200	1600	MLM_2063	2022685	2023435	short-chain dehydrogenase	Pseudogene
		MLM_2065	2023485	2024215	putative short-chain dehydrogenase	Pseudogene
2707801- 2721800	14000	MLM_2691	2707807	2708890	mobile element protein	Pseudogene
		MLM_2693	2708891	2709731	aldo-keto reductase	Pseudogene
		MLM_2695	2709960	2710996	putative oxidoreductase YncB	Pseudogene
		MLM_2696	2711067	2711455	putative LysR-family transcriptional regulator	Pseudogene
		MLM_2697	2711736	2712220	putative LysR-family transcriptional regulator	Pseudogene
		MLM_2698	2712184	2712453	TetR family transcriptional regulator	Pseudogene
		MLM_2699	2712541	2713507	quinone oxidoreductase	Pseudogene
		MLM_2701	2713577	2714488	putative LysR-family transcriptional regulator	Protein-coding
		MLM_2702	2714586	2715407	hydroxymethylglutaryl-CoA lyase	Protein-coding
		MLM_2703	2715424	2715951	acyl dehydratase	Protein-coding
		MLM_2704	2715951	2717291	4-hydroxybutyrate:acetyl-CoA CoA transferase	Protein-coding
		MLM_2705	2717306	2719057	pyruvate oxidase	Protein-coding
		MLM_2706	2719437	2719946	nitroreductase	Protein-coding
		MLM_2707	2720165	2720832	short-chain dehydrogenase/reductase	Pseudogene
3121401- 3129500	8100	MLM_3070	3121408	3121638	uncharacterized protein	Pseudogene
		MLM_3071	3122087	3123147	uncharacterized protein	Pseudogene
		MLM_3074	3123144	3124244	uncharacterized protein	Pseudogene
		MLM_3077	3124699	3125469	DUF58 domain-containing protein	Pseudogene
		MLM_3078	3125469	3125887	mobile element protein	Pseudogene
		MLM_3079	3126127	3127092	moxR-like ATPases	Pseudogene
		MLM_3080	3127190	3129141	uncharacterized protein	Pseudogene

Table S3: Genes implicated in the biosynthesis of GPLs in the *Mycobacterium avium* complex.

Gene name(s)	Product and/or function	*M. avium* 104	*M. colombiense* CECT_3035	*M. intracellulare* ATCC_13950	*M. lepraemurium* Hawaii
mmps4	Required for assembly of GPL synthesis enzymes in the cell membrane	MAV_3247	MCOL_V218616	OCU_30810	MLM_2607
mmpL4a	Required for assembly of GPL biosynthases in the cell membrane	MAV_3248	MCOL_V218621	OCU_30820	MLM_2608
mmpL4b	Required for assembly of GPL biosynthases in the cell membrane	MAV_3249	MCOL_V218626	OCU_30830	MLM_2609
MAV_3362	Unknown function	MAV_3362	MCOL_V219166	OCU_32010	MLM_1724
rmlA	Glucose-1-phosphate thymidylyltransferase	MAV_4820	MCOL_V201710	OCU_46990	MLM_3982
rmlB	NAD dependent epimerase/dehydratase family protein	MAV_3269	MCOL_V218756	OCU_31110	MLM_2632
mtfA	3-O-methyltransferase	MAV_3268	WP_075235767.1*	OCU_27880	MLM_2631 (pseudogene)
mtfB	Rhamnose 4-O-methyltransferase	MAV_3266	MCOL_V218751	OCU_31100	MLM_2630
mtfC	Rhamnose 4-O-methyltransferase	MAV_3261	MCOL_V218736	OCU_31070	MLM_2626
mtfD	Rhamnose 3-O-methyltransferase	MAV_3260	MCOL_V218731	OCU_31060	MLM_2625
gtfA	Glycosyl transferase family protein	MAV_3265	MCOL_V218746	OCU_31090	MLM_2629
gtfB	Glycosyl transferase family protein	MAV_3258	WP_065074816.1**	OCU_31040	MLM_2623
gtfD	Putative glycosyl transferase	MAV_3253	WP_075237225.1*	OCU_30870	Not found
rtfA	Putative glycosyl transferase	MAV_3262	MCOL_V218741	OCU_31080	MLM_2627
atf	Putative acyltransferase	MAV_3274	MCOL_V218771	OCU_31140	MLM_2635 (pseudogene)
gplH	Unknown function	MAV_3245	MCOL_V218611	OCU_30790	MLM_2606
pstA (mps1)	Linear gramicidin synthetase subunit D	MAV_3244	MCOL_V218606	OCU_30780	MLM_2605
pstB (mps2)	Linear gramicidin synthetase subunit B	MAV_3243	MCOL_V218601	OCU_30770	MLM_2604
gap	Required for GPL export.	MAV_3059	MCOL_V222543	OCU_28790	MLM_2430 (pseudogene) + MLM_2431 (pseudogene)
sap	Dgpf domain family protein	MAV_4518	MCOL_V200150	OCU_43890	MLM_3677
ecf	RNA polymerase ECF-subfamily protein sigma factor	MAV_4519	MCOL_V200155	OCU_43900	MLM_3678
fadE5	Putative acyl-CoA dehydrogenase	MAV_3309	MCOL_V218931	OCU_31500	MLM_1767 (pseudogene)
MAV_2461	Uncharacterized protein	MAV_2461	MCOL_V208010	OCU_24450	MLM_2147 (Partial hit. Also, the ORF is much longer)
pks	Polyketide synthases	MAV_1763	MCOL_V205565	OCU_18380	MLM_2746
papA3	Condensation domain protein	MAV_1762	MCOL_V205560	OCU_18360	MLM_2749
mmpL10	Transport protein	MAV_1761	MCOL_V205550	OCU_18350	MLM_2750
fadD23	acyl-CoA synthase	MAV_1759	WP_081293032.1*	OCU_18330	MLM_2752
pe	Uncharacterized protein	MAV_1760	WP_065054842.1*	OCU_18340	MLM_2751
MAV_1758	Uncharacterized protein	MAV_1758	WP_076099905.1*	OCU_12310	MLM_2753
dhgA	Dehydrogenase DhgA	MAV_3259	Not found	OCU_31050	MLM_2624

*As in *M. avium* 104
**The gene is absent in *M. colombiense* CECT_3035, but is present in other *M. colombiense* strains. A representative protein is given.
***The gene is present in *M. colombiense* CECT_3035, but is not annotated. A representative protein from another strain is given.

Chapter 3. : Evolution of *M. leprae*

Ancient *Mycobacterium leprae* genomes reveal an unexpected diversity of leprosy in medieval Europe

Verena J. Schuenemann[1,2,3]*, Charlotte Avanzi[4]*, Ben Krause-Kyora[5,6]*, Alexander Seitz[7]*, Alexander Herbig[6], Sarah Inskip[8], Marion Bonazzi[5], Ella Reiter[1], Christian Urban[1], Dorthe Dangvard Pedersen[9], G. Michael Taylor[10], Pushpendra Singh[4,11], Graham T. Stewart[10], Petr Velemínský[12], Jakub Likovsky[13], Antónia Marcsik[14], Erika Molnár[14], Gyorgy Pálfi[14], Valentina Mariotti[15,16], Alessandro Riga[17], M. Giovanna Belcastro[15,16], Jesper L. Boldsen[9], Almut Nebel[5], Simon Mays[18], Helen D. Donoghue[19], Sonia Zakrzewski[20], Andrej Benjak[4], Kay Nieselt[7#], Stewart T. Cole[4#]; Johannes Krause[1,2,6#].

1. Institute for Archaeological Sciences, University of Tübingen, 72070 Tübingen, Germany.
2. Senckenberg Centre for Human Evolution and Palaeoenvironment, University of Tübingen, 72070 Tübingen, Germany.
3. Institute for Evolutionary Medicine, University of Zurich, 8057 Zurich, Switzerland.
4. Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
5. Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany.
6. Max Planck Institute for the Science of Human History, 07745 Jena, Germany.
7. Centre for Bioinformatics, University of Tübingen, 72076 Tübingen, Germany.
8. McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge, CB2 3ER, UK.
9. Department of Anthropology (ADBOU), Institute of Forensic Medicine, University of Southern Denmark, Odense S, Denmark.
10. Department of Microbial Sciences, Faculty of Health and Medical Sciences, University of Surrey, GU2 7XH Guildford, UK.
11. Department of Microbiology and Biotechnology Centre, The Maharaja Sayajirao University of Baroda, Vadodara 390002, India (current address).
12. Department of Anthropology, National Museum, Prague, Czech Republic.
13. Department of Archaeology of Landscape and Archaeobiology, Institute of Archaeology, Academy of Sciences of the Czech Republic, Prague, Czech Republic.
14. Department of Biological Anthropology, University of Szeged, Hungary.
15. Laboratorio di Bioarcheologia e Osteologia Forense, Dipartimento di Scienze Biologiche, Geologiche e Ambientali, Via Selmi 3, 40126 Bologna, Italy.
16. ADÉS, UMR 7268 CNRS / Aix-Marseille Université/EFS, Aix-Marseille Université, Marseille, France.
17. Dipartimento di Scienze della vita e biotecnologie, Via Luigi Borsari 46, 44121 Ferrara, Italy.
18. Historic England, Portsmouth, UK.
19. Centre for Clinical Microbiology, Division of Infection and Immunity, University College London, London NW3 2PF, UK.
20. Department of Archaeology, University of Southampton, Southampton, UK.

* These authors contributed equally to this work
† Corresponding author

PNTD, under review

Contributions: study design, skeletal sampling, ancient DNA experiments, interpretation of results, manuscript preparation

Abstract

Studying ancient DNA allows us to retrace the evolutionary history of human pathogens, such as *Mycobacterium leprae*, the main causative agent of leprosy. Leprosy is one of the oldest recorded and most stigmatizing diseases in human history. The disease was prevalent in Europe until the 16th century and is still endemic in many countries with over 200,000 new cases reported annually.

Previous worldwide studies on modern and European medieval *M. leprae* genomes revealed that they cluster into five separate lineages of which two were present in medieval northwestern Europe. In this study, we analyzed 10 new medieval *M. leprae* genomes including the so far oldest *M. leprae* genome from one of the earliest known cases of leprosy in the United Kingdom—a skeleton from the Great Chesterford cemetery with a calibrated age of 415–545 AD. This dataset provides a genetic time transect of *M. leprae* diversity in Europe over the past 1500 years. We find four of the five known distinct *M. leprae* lineages to be present in the early medieval period, and three lineages were detected within a single cemetery from the high medieval period. Altogether these findings suggest a higher genetic diversity of *M. leprae* strains in medieval Europe at various time points than previously assumed. The resulting more complex picture of the past phylogeography of leprosy in Europe impacts current phylogeographical models of *M. leprae* lineages. It suggests alternative models for the past spread of leprosy such as a wide spread prevalence of different lineages in Eurasia already in Antiquity or maybe even an origin in Western Eurasia. Furthermore, these results highlight how studying ancient *M. leprae* strains improves understanding of the history of leprosy worldwide.

Author's Summary

Many mysteries surround leprosy, which is one of the oldest recorded diseases of humankind. The origin and past spread of its main causative agent, *Mycobacterium leprae*, remain unknown although many attempts have been made to reconstruct its past from historical and archeological sources. Analysis of ancient *M. leprae* genomes reconstructed from archaeological remains can contribute greatly to reconstructing the origin and evolution of this pathogen. With a new set of ancient *M. leprae* genomes from Europe, we traced back a so far unrecognized past diversity, which places Europe as a key region for the early spread and worldwide dissemination of leprosy. Our results hint to the potential dynamic changes in the prevalence of different *M. leprae* strains in Europe during Antiquity, and highlight the need to study ancient pathogen genomes in order to better understand our past.

Introduction

Leprosy resulting from the infection with *Mycobacterium leprae* has been prevalent since early history. Signs of its existence occur in historical texts [1, 2] as well as in the osteological and archeological records [3-5]. Widespread in medieval Europe, and peaking between the 12th and 14th century, leprosy declined in the 16th century and subsequently disappeared from Europe [6, 7]. Nowadays *M. leprae* is prevalent worldwide except in Europe and is genetically represented by five distinct lineages with different geographic distributions. Lineage 0 is mainly found in Eastern Asia (Japan, China, New Caledonia) whereas lineage 1 is mostly detected in Southern and Eastern Asia (Thailand, India, Southern Japan)[8-12]. Lineage 2 is reported from the Near East [8], however, no such modern day example (genome) has been sequenced so far. Lineage 3 is present in Latin America and recently spreading in the southwestern USA nine-banded armadillo population [13] with occasional zoonotic transmission to humans [14]. Another non-human *M. leprae* reservoir, also harboring a lineage-3 strain, is the red squirrel in England [15]. Finally, lineage 4 is present in West Africa and South America [8, 9, 16]. *M. leprae* lineages correspond to specific single nucleotide polymorphism (SNP) types or subtypes [9], a nomenclature widely used in comparative genomics, consisting of four SNP types (1-4) and 16 SNP subtypes (A to P) [8]. While SNP types are based on a limited amount of SNPs [8], *M. leprae* lineages include full genomes and therefore reflect more complexity [9].

Investigations on the evolutionary history of *M. leprae* have elucidated on the past phylogeography and diversity of the leprosy bacillus in Europe. Recently sequenced medieval *M. leprae* genomes reveal the

presence of at least two distinct *M. leprae* lineages in medieval Northwestern Europe [9]. Furthermore, the data indicate a high level of genetic conservation during the last 1000 years. There appears to be a close relationship of a group of late medieval strains with contemporary strains present today in the Southwestern USA [9] infecting humans and armadillos [13] as well as red squirrels in England [15]. Two medieval genomes from a cemetery in the UK suggested a possible predominance of lineage 2 during the 10th to the 12th century in Northwestern Europe, while lineage 3 was more frequent during the late medieval era [17]. However, the past diversity and population structure of *M. leprae* at different time points in other parts of Europe still remain unclear.

To address these key questions, we sequenced four whole genomes of *M. leprae* strains identified in early medieval leprosy cases from various parts of Europe -Italy, Hungary, Czech Republic and the UKincluding one of the oldest leprosy cases in the UK from an early Anglo-Saxon cemetery in Great Chesterford, radiocarbon dated to AD415-545 [18]. Furthermore, we sequenced six genomes from the Odense St. Jørgen cemetery in Denmark to assess the diversity of *M. leprae* in one location at a particular point in time. Our results reveal a high diversity of *M. leprae* lineages in early medieval Europe, where four of five known *M. leprae* lineages were identified. Furthermore, three lineages of *M. leprae* were found within the St. Jørgen cemetery alone, indicating a high level of strain diversity in medieval Europe.

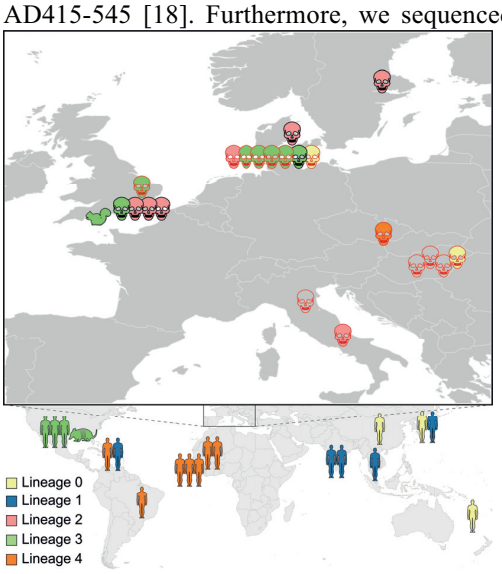


Figure 1: Worldwide distribution of the ancient and modern *M. leprae* strains analysed in this study. Skulls represent strains from osteological specimens dated to the medieval area. Human silhouettes are modern strains; squirrel and armadillo silhouettes are strains from the red squirrel and the nine-banded armadillo. Skulls outlined in red are the new *M. leprae* genomes reconstructed in this study, while skulls outlined in black represent previously sequenced ancient genomes. Grey skulls (outlined in red) are leprosy samples from this study that did not yield sufficient sequence for analysis. The main *M. leprae* lineages (see Figure 2) are color-coded.

Results

Sample collection and screening. The potential leprosy cases identified based on characteristic skeletal deformations that were analyzed in this study span different European countries and time periods (Figure 1 and Table S1 for details of all samples except the negative Danish ones): two from Italy (4th BC and 7th AD) [19-21], one from UK (5th to 6th century) [18], four cases from Hungary (7th to 11th century) [20], one from the territory of today's Czech Republic (9th to 12th century) [20], and 87 from Denmark (11th-14th century). The screening for the Danish samples was carried out with PCR [9] and direct shotgun sequencing of double-stranded Illumina libraries [22, 23] followed by metagenomics analysis using MALT [24]. Samples with more than 1% (n=6, Table S1) of all raw reads mapping to the *M. leprae* TN reference genome (RefSeq ID NC_002677.1) were used for whole genome sequencing by shotgun sequencing without any prior enrichment. For all other samples double-stranded Illumina libraries [22, 23] were created and screened for *M. leprae* DNA using a bead capture approach [25] on three genomic loci, *gyrA, proS* and *RLEP*, as detailed previously [9], yielding between 0.4x and 700x-fold average coverage for the enriched genes. DNA misincorporation patterns characteristic for ancient DNA were calculated to assess the authenticity of the retrieved DNA [26, 27]. As observed previously [9], ancient *M. leprae* DNA contained a lower percentage of misincorporation patterns, between 11 and 21% (Table S1), compared to what is expected for human DNA of the same region and age [28]. This consistency underlines the exceptionally good long-term preservation of *M. leprae* DNA within cells as already commented previously [9].

Table 1: Results of the genome-wide analysis for the samples with sufficient coverage

Sample Name	Mapped non-duplicate reads	Mean Coverage	Genome fraction covered by at least 5 non-duplicate reads	average fragment length	SNPs per sample
Jorgen404	394594	10.8	98.97	89.65	87
Jorgen427	1087657	29.8	99.98	89.65	54
Jorgen507	1997693	59.8	99.99	97.84	183
Jorgen533	378098	11.1	98.07	95.61	108
Jorgen722	385329	11.2	98.44	94.56	118
Jorgen749	1353305	40.1	99.99	96.92	91
GC96	777067	20.5	97.38	86.21	96
T18	329362	7	61.32	69.59	25
Body188	3561241	116	99.99	106.45	101
SK11	442725	7.8	68.99	57.51	45

Genome-wide enrichment, sequencing, and analysis. On the basis of the screening results (Table S1), samples containing *M. leprae* DNA were subjected to whole-genome enrichment and sequencing. The libraries were prepared using an enzymatic DNA damage repair [29]. Prior to sequencing, all libraries were enriched using an array spanning the *M. leprae* genome [9, 30], except for the Danish samples, which were directly shotgun sequenced. Between 406,241 and 12,227,587 short reads were mapped to the *M. leprae* TN reference genome (RefSeq ID NC_002677.1) using the EAGER pipeline [31] and all samples with at least 7x-fold mean coverage were selected for further analysis (Tables 1, S2). For the resulting 10 samples, we used the Genome Analysis Toolkit (GATK) to generate a mapping assembly to call reference bases and variants from the mapping.

Phylogenetic analysis. We reconstructed the phylogeny for the 10 newly sequenced and well-resolved medieval genomes together with the previously sequenced medieval [9, 17] and modern genomes [9, 10, 13-16, 32], from a total of 1071 SNP positions. The four early medieval genomes fall on different branches in the maximum parsimony tree (Figure 2A). The oldest *M. leprae* genome from Great Chesterford (GC96) belongs to lineage 3, as the armadillo, red squirrel, and other medieval and modern human strains [9, 15]. Interestingly, the *M. leprae* strains isolated from red squirrels from England [15] are placed tightly between GC96 from Great Chesterford (5^{th}-6^{th} century) and SK2 from Winchester (10^{th}-11^{th} century) (Figure 2).

The medieval genome from Italy (T18) falls within lineage 2, whereas SK11 from Hungary clusters with lineage 0 strains. The Body188 strain from Czech Republic belongs to lineage 4 and is ancestral to contemporary strains from Western Africa. The six individuals from the St. Jørgen cemetery in Denmark, which was established 1270 and existed to 1560, are 14C-dated with overlapping periods and mean values from the 12^{th} to the 14^{th} century (Table S1), a period in which the majority of burials took place [33]. The six *M. leprae* genomes obtained from these individual cluster within three different lineages (Figure 2A): lineage 0 (n=1), lineage 2 (n=1) and lineage 3 (n=4).

Table 2: Time to the most recent common ancestor (tMRCAs) for the entire *M. leprae* tree and individual lineages

lineage	constant pop. size		Bayesian Skyline	
	mean tMRCA [y]	95% HPD [y]	mean tMRCA [y]	95% HPD [y]
whole tree	5162	4012-6393	4946	3868-6147
0	4745	3620-5951	4541	3491-5752
1	2488	1925-3084	2401	1880-2984
2	1830	1558-2136	1836	1575-2139
3	2395	1990-2849	2338	1958-2785
4	1771	1383-2225	1774	1388-2206

SNP analysis. We analyzed the 1071 SNP positions identified in the whole dataset for the effects that they might have on particular genes (Table S3). No polymorphisms specific only to all ancient strains were found. Non-synonymous SNPs (nsSNPs) specific to each lineage are given in Table S4.

BEAST analysis. Using the radiocarbon dates for the ancient samples and the isolation dates for the modern samples as tip calibration points we estimated the divergence time for all *M. leprae* strains using BEAST [34] assuming a constant population size [35] and a variable population size by application of a Bayesian Skyline model [36]. The age of the most recent common ancestor (MRCA) was estimated to 5,162 y BP (4,012–6,393 y 95% Highest Posterior Density (HPD)) under the constant population size model, and 4,946 y BP (3,868–6,147 y 95% HPD) assuming the variable population size (Bayesian Skyline) model (Figure 2B, Table 2). The estimated mean age of the MRCA is about 1,500 y older compared to the dating analysis previously performed [9] with fewer modern day and medieval genomes available. The shift was mainly due to the ancient genomes from lineage 0, which show a higher variability, compared to other lineages. However, the 95% HPD intervals of both analyses overlap, increasing the confidence of the estimates. The age of the MRCAs of the main lineages varied between 4,745 y BP (3,620–5,951 y 95% HPD) for lineage 0 and 1,771 y BP (1,383 – 2,225 y 95% HPD) for lineage 4 (Table 2). The mutation rate for *M. leprae* was estimated to 6.05 x 10^{-9} (4.69 – 7.51 x 10^{-9} 95% HPD) and 6.25 x 10^{-9} (4.77 – 7.69 x 10^{-9} 95% HPD) substitutions per site per year for constant and variable population size, respectively. This rate can be also depicted as 19-24 mutations per 1000 years and genome.

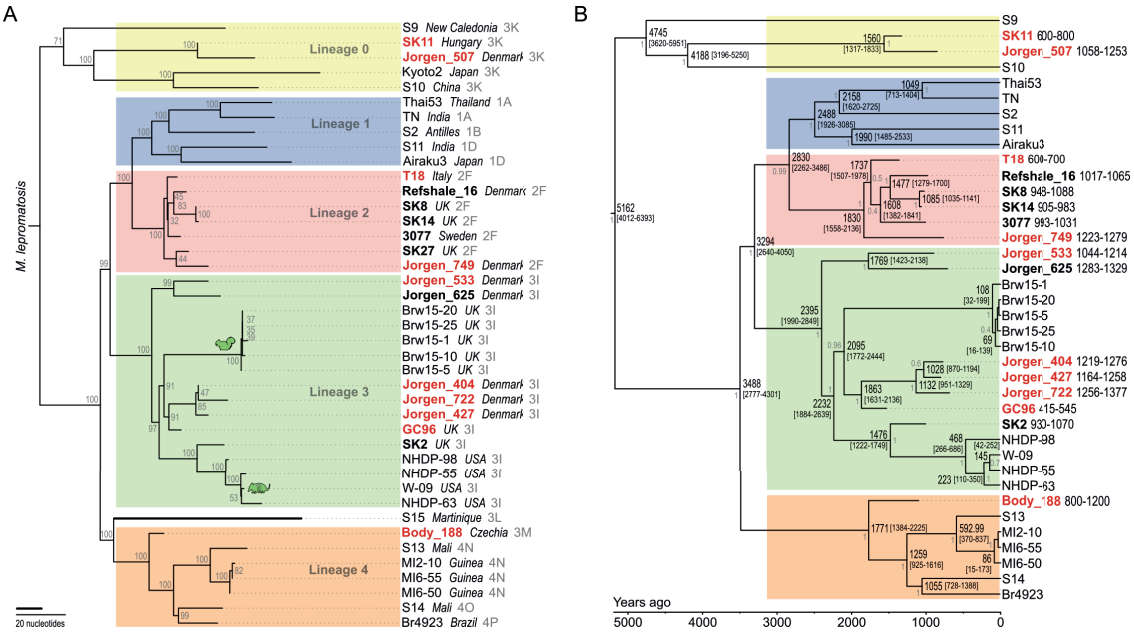


Figure 2: Phylogenetic analysis of ancient and modern *M. leprae* strains. (A) Maximum parsimony tree reconstructed from 1071 informative SNP positions. The tree is drawn to scale, with branch lengths representing number of substitutions. Note that the exceptionally long branch of S15 (thicker line) was reduced to 50% to save space. *M. lepromatosis* was used as outgroup. Tip labels for each sample show the name, the country of origin (italic) and the SNP subtype (grey). Names of strains are in bold, with the novel strains from this study in red. Animal symbols indicate strains isolated from red squirrels or armadillos. Bootstrap values (500 replicates) are shown next to each node. (B) Bayesian phylogenetic tree calculated with BEAST 1.8.1. Median divergence times are shown on each node in years before present (the 95% Highest Posterior Density ranges are given in square brackets). Posterior probabilities for each node are shown in grey.

Discussion

The 10 newly sequenced medieval *M. leprae* genomes allow us to trace back the last 1500 years of leprosy history in Europe and illustrate the high diversity of lineages circulating in Europe during this time transect.

Previous genome wide data suggested the existence of only two lineages in medieval Europe, lineage 2 present at least from the 10th to the 12th century and lineage 3 in the late medieval era [9, 17]. Our results reveal a higher diversity in medieval Europe than anticipated. Strains belonging to the ancestral lineage 0 and associated with modern strains from East Asia were found in Hungary (SK11) and Denmark (Jorgen 507) confirming the assumed presence of this lineage in Europe since at least the 7-8th century [20]. The so far oldest *M. leprae* genome from Great Chesterford falls into lineage 3 suggesting the existence of this lineage in Europe at least since the 5th century. With the T18 strain we can now also detect lineage 2 in early medieval Central Italy, which until now was found only in Northern Europe [9, 17]. The Body188 strain from Czech Republic on the other hand is ancestral to modern lineage 4 strains from West Africa and Brazil. This strain belongs to the SNP subtype 3M, which has not yet been sequenced and is rarely identified in modern samples [8]. Based on our phylogenetic tree we can now group SNP subtype 3M together with modern SNP-type 4 strains into lineage 4. Body188 provides therefore a link between Europe and West Africa, where the contemporary SNP-type 4 is predominant [8]. However, a higher resolution with more ancient and modern strains is needed before we can understand this link.

The dynamics of leprosy transmission throughout human history is not fully resolved, but characterization and geographic association of the most ancestral strains are crucial for deciphering leprosy's origin, which still remains elusive. This is in part due to the scarcity of convincing evidence of leprosy in historical records that predate the Common Era [37]. The earliest accepted written record of leprosy is in the Sushruta Samhita, an old Indian text on medicine and surgery dated around 600 BC [1], and with the exception of a limited amount of potential cases, such as the so far oldest one from India dated around 2000 BC [3] or potential cases from Italy and Hungary dated to the 4th-3rd century BC [19, 38]. However, all are not yet confirmed on the molecular level. The oldest osteological cases of leprosy that could be detected by molecular methods are from around the beginning of the Common Era [8, 20], when written records also become more abundant [2, 37]. Therefore, at present, we are limited to molecular methods to decipher the time of *M. leprae*'s origin and its early spread in humans. Having more ancient genomes in a dating analysis should result in more reliable estimates. In this study, every major *M. leprae* lineage had one or more representative ancient genomes present. This is the likely reason why the estimated oldest age of the MRCA of *M. leprae* is around 1,500 years older compared to the earlier calculations with fewer ancient genomes [9]. The new dating results justify searching for even older osteological cases of leprosy than currently available using well-established methods for identification of potential cases [39].

The high diversity of *M. leprae* in medieval Europe, spanning almost all its major lineages including the most basal ones in the phylogeny, has not been found anywhere else worldwide. Based on this observation two antipodal models can be developed: an origin of leprosy in Western Eurasia, maybe even in Europe, and spread from there into the rest of the world (model 1), and the introduction of diverse *M. leprae* lineages to Europe from different regions in the world during and before the Medieval era (model 2) that likely displayed a high diversity of *M. leprae* strains. The main argument supporting both models is that Europe was an important socioeconomic hub from Classical Antiquity onwards. The dynamic geopolitical changes, wars and the well-established trading routes contributed to continuous contacts with the neighboring regions and migrations within and outside Europe and allow fast and multiple exchanges of pathogens in both directions. In favor of model 1 (Western Eurasian origin) is the argument that most medieval European *M. leprae* strains are ancestral to modern strains (Figure 2). On the other hand, supporting model 2 (multiple introductions into Europe) is the prevalence of the most ancestral lineage 0 (subtype 3k) in modern day China, Japan and Korea [8, 40-42] and in the Middle East [8]. It has to be noted that strains from Central Asia are poorly analyzed to date. Overall, the current genomic data convincingly suggest Eurasia as a broad area of origin and of the early spread of *M. leprae*. Including more ancient strains from different parts of the world in a phylogenetic context may enable us to narrow down this area. This is especially important in areas where past events have

become blurred by extensive population mixing, bottlenecks or strain replacements.

In summary, our results provide a genetic time transect for the diversity of *M. leprae* within the medieval period in Europe and allow us to gain a better understanding of the past phylogeography of this pathogen. Discovery of other ancient *M. leprae* strains, especially from Asia will provide new details on the diversity of *M. leprae* in the past and help develop models for its global spread.

Methods

Ancient DNA extraction and library preparation DNA extractions from the samples JK3187 to JK3195, GC96F and GC96C (Table S1) were conducted from 30-50 mg bone powder for each sample in clean room facilities dedicated to ancient DNA work at the University of Tübingen. The Danish samples were processed separately at the University of Kiel following the same protocols as described below for the samples processed in Tübingen. A silica purification protocol was applied as previously described [43] using the following modifications: the Zymo-Spin V columns (Zymo Research) were UV irradiated for 60 minutes and the total elution volume was raised to 100 µl.

Aliquots of 20 µl from JK3187 to JK3195, GC96F and GC96C were converted into double-stranded Illumina libraries [22]. The adapter-ligated fragments were quantified through a quantification assay using the primers IS7 and IS8 [22], the DyNAmo Flash SYBR Green qPCR Kit (Biozym) and the Lightcycler 96 (Roche).

Following established protocols [22, 23] sample-specific indexes were added in the next step to both library adapters via amplification to create double indexed libraries. Extraction and library blanks were treated accordingly. These libraries were used in the following for initial screening approaches.

For genome-wide enrichment and sequencing additional libraries were prepared from 30 to 50 µl aliquots of all DNA extracts according to the methods described above [22, 23] with one modification: One additional stepthe treatment of all extracts and blanks with uracil-DNA glycosylase (UDG) and endonuclease VIIIwas included into library preparation to avoid potential sequencing artefacts caused by the characteristic ancient DNA damage profile produced by the deamination of cytosine to uracil over time [29].

For all indexed libraries a subsequent amplification was performed as detailed in Schuenemann and colleagues [9].

Enrichment and sequencing for sample screening All samples except the Danish ones were screened for *M. leprae* preservation. Three *M. leprae* genes ML0006 (*gyrA*), ML1553 (*proS*) and RLEPwere selected as targets for DNA enrichment and converted into bait DNA using Long Range PCR products as described previously [9]. Following the bead enrichment protocol by Maricic and colleagues [25] a hybridization of the amplified libraries, pooled in an equimolar amount, to the DNA bait was carried out.

Subsequently to the bead enrichment the libraries were sequenced on an Illumina HiSeq 2500 platform using a paired-end dual index run with 2*101+8+8 cycles (for the samples GC96F and GC96C) and for 2*125+8+8 cycles (for the samples JK3187 to JK3195) following the manufacturer's protocols for multiplex sequencing (TruSeq PE Cluster Kit v3-cBot-HS).

The screening for the Danish samples was carried out with PCR for parts of specific genes (18kDa antigenic protein gene and repetitive element RLEP gene)[9] and direct shotgun sequencing of double-stranded Illumina libraries [22, 23]. The sequencing for the shotgun data was performed on the Illumina HiSeq 4000 platform at the Institute of Clinical Molecular Biology, Kiel University, by 2×75 cycles using the HiSeq v4 chemistry and the manufacturer's protocol for multiplex sequencing.

Genome-wide enrichment and sequencing The UDG treated libraries (one from JK3187 to JK3195, GC96F and two from GC96C) were enriched genome-wide with two rounds of hybridization capture following the protocol detailed before [44]. The design of the 1 million Agilent SureSelect arrays used in the study was described previously [9]. In a first approach equimolar pools of GC96F and GC96C1 were enriched on one array, in a second experiment equimolar pools GC96C1 and GC96C2 on two arrays. In a third approach the samples JK3187 to JK3189 and JK3192 to JK3195 were enriched on one array and JK3190 and JK3191 on a separate one. In all three experiments the extraction and library blanks were enriched separately on an additional array. After the first round of hybridization, captured products were eluted in 490 µl H_2O and quantified via a quantitative PCR with the IS5 and IS6 primer set [22], the DyNAmo Flash SYBR Green qPCR Kit (Biozym) and the Lightcycler 96 (Roche). In a subsequent amplification the eluted products were amplified for 17 to 20 cycles in 100 µl reactions using 24 µl template, 4 units of AccuPrime Pfx DNA polymerase (Invitrogen), 1 unit of 10× AccuPrime buffer (containing dNTPs) and 0.3 µM of the primers IS5 and IS6 [22] and the following thermal profile: a 2-min initial denaturation at 95°C, 17 to 20 cycles consisting of 15 sec denaturation at 95°C, a 30-sec annealing at 60°C and a 2-min elongation at 68°C, followed by a 5-min final elongation at 68°C. The amplified enriched library pools were purified using MinElute columns (Qiagen) following the manufacturer's protocol and quantified via an Agilent 2100 Bioanalyzer DNA 1000 chip. All pools were then enriched in a second round of hybridization capture using the same number of arrays as in the first round. After the second round the capture products were eluted and subsequently processed as previously described with the following modification: 48 µl template was used per amplification reaction.

After the enrichment paired-end dual indexing sequencing was performed on an Illumina HiSeq 2500 platform using 2*101+8+8 cycles (for the samples GC96F and GC96C) and for 2*125+8+8 cycles (for the samples JK3187 to JK3195) using the manufacturer's protocols for multiplex sequencing (TruSeq PE Cluster Kit v3-cBot-HS). A second round of sequencing was conducted for the samples JK3187 to JK3195 on an Illumina HiSeq 4000 platform using a single-end dual index run with 75+8+8 cycles.

Genome-wide sequencing for the Danish samples The sequencing of the six Danish samples was carried out on the Illumina HiSeq 4000 platform at the Institute of Clinical Molecular Biology, Kiel University, by 2×75 cycles using the HiSeq v4 chemistry and the manufacturer's protocol for multiplex sequencing.

Data processing for screening analysis The data processing after screening for all samples was carried out as previously described [9] with modifications of using the EAGER pipeline [31]. For all samples except the Danish ones the processed reads were mapped to the three *M. leprae* loci *gyrA*, *proS*, and *RLEP* and characteristic damage profiles were calculated for the *M. leprae* DNA to assess the authenticity of the ancient DNA [26, 27].
For the Danish samples a metagenomics analysis using MALT [24] was conducted after the processing through the EAGER pipeline [31]. In addition, the shotgun data was also mapped to the *M. leprae* TN reference genome (RefSeq ID NC_002677.1).

Data processing for genome-wide analysis The sequenced reads from all samples subjected to genome-wide enrichment were analyzed with the EAGER pipeline [31].
Read preprocessing of sequenced genome samples. The first step of the pipeline was adapter clipping, read merging and subsequent quality trimming using the tool "Clip&Merge". For all properly merged reads, only the merged consensus sequence was used for the subsequent mapping steps. On average about 89% of all paired-end reads were merged in each sample (see Table S2 for detailed results of all samples). For those read pairs that could not be merged because the overlap region was shorter than 10 nucleotides or for which the corresponding read was removed during the combined adapter clipping and quality filtering step, the respective single-end reads were first trimmed at the 3'end such that all bases have a Phred quality score of at least 20 and then mapped individually.

Mapping. After adapter clipping, merging and quality trimming, the resulting reads for all samples were mapped using the *M. leprae* TN genome (RefSeq ID NC_002677.1) as a reference. All reads (merged and unmerged) were treated as single-end reads and mapping was performed using BWA [45] aln/samse subcommands, with an error rate (-n) of 0.2 to assure high specificity. The PCR duplicates were removed with MarkDuplicates from the Picard tools (https://broadinstitute.github.io/picard). The mapping was evaluated with QualiMap [46].
Mapping assembly. After mapping and duplicate removal, the Genome Analysis Toolkit (GATK) [47] was used to generate a mapping assembly for each sample that had at least 60% genome coverage and a minimum of 5 reads per base. For this procedure the UnifiedGenotyper module of GATK following the GATK Best Practice's Guidelines was applied to call reference bases and variants from the mapping. The reference base was called if the genotype quality of the call was at least 30, the position was covered by at least 5 reads and at least 90% of the bases at this position agreed with the reference. A variant position (SNP) was called if the following criteria were met: i) the position was covered by at least 5 reads; ii) the genotype quality of the call was at least 30 and III) the minimum SNP allele frequency was 90%. If neither of the requirements of a reference base call nor the requirements for a variant call were met, the character 'N' was inserted at the respective position. To keep the potential introduction of too many 'N' characters as low as possible in the case of low coverage genomes, in cases where a position had a coverage between 5-9 reads, the major allele was called if it was found in all but 1 read. For the generation of draft genome sequences we used the VCF2Genome tool, also available in EAGER.

Processing of published modern samples. The reads for the modern samples *S2, S10, S11, S13, S14*, and *S15,* which were previously published [9], were single-ended. Thus these samples were not merged but only adapter-clipped and quality trimmed. Afterwards the reads were treated exactly the same as the other samples. Furthermore a correction of the country of origin for strain S15 should be added in this context: Strain S15 corresponds to strain 92041 [48] and it was isolated from a lepromatous leprosy patient originally from Martinique. The origin of S15 was erroneously attributed to New Caledonia in Monot et al., 2009 (Monot et al., 2009, Supplementary Table S3) and the error was subsequently propagated in several publications [9, 10, 15, 16].

Processing of published genomes. In order to apply our analysis pipeline also to those samples for which complete genomic sequences are available in GenBank (*Br4923*, and *TN*) and *Kyoto2* (Matsuoka personal communication), we produced artificial reads using an in-house tool (Genome2Reads). In a tiling approach, we produced reads of length 150 nucleotides with a tiling offset of 2, resulting in an average genome coverage of 75X. For the resulting samples we applied the same mapping, SNP calling and genome reconstruction procedure as for the sequenced samples in order to obtain consistent and comparable results. The genome sequence of *Mycobacterium lepromatosis* (GenBank JRPY00000000.1), used as the outgroup for the Maximum parsimony tree, was aligned against *M. leprae* using LAST with the gamma-centroid option [49].
Phylogenetic analyses. For the phylogenetic analysis, a SNP alignment based on 1071 informative positions was generated. This alignment contained all positions where a SNP was called in at least one sample. Positions covered in a negative control [9] were excluded from subsequent phylogenetic analyses. From the resulting alignment of length 1071 bp a phylogenetic tree was created with MEGA [50] using the Maximum parsimony method, with partial deletion on a site coverage cutoff of 80% and 500 bootstraps.

Dating analysis We estimated divergence times and substitution rates by application of the Bayesian framework BEAST 1.8.1[36]. In this analysis we included all ancient and modern strains (Figure 2) except for strain S15 because of its extraordinary branch length, which is probably due to selection pressure from anti-leprosy treatment [9], strain SK27 because of its low coverage and strain Kyoto2 because raw data was not available and quality assessment was therefore not possible. We performed two analyses using a

constant population size coalescent prior [35] and a Baysian Skyline model [36] for variable population size, respectively. For both analyses we applied a lognormal relaxed clock and an HKY substitution model. For ancient strains tip dates were uniformly sampled from dating intervals [9, 17, 18, 20] (samples from this study Table S1) whereas for modern strains tip dates were set as isolation dates [8-10, 13-16]. For each model an MCMC run was carried out with 300,000,000 iterations discarding the first 30,000,000 iterations as burn-in.

References:

1. Dharmendra. Leprosy in ancient Indian medicine. International Journal of Leprosy. 1947;15:424–30.

2. Hulse EV. The nature of biblical "leprosy" and the use of alternative medical terms in modern translations of the Bible. The Palestine Exploration Quarterly. 1975;107:87-105.

3. Robbins G, Tripathy VM, Misra VN, Mohanty RK, Shinde VS, Gray KM, et al. Ancient skeletal evidence for leprosy in India (2000 B.C.). PLoS One. 2009;4(5):e5669. Epub 2009/05/30. doi: 10.1371/journal.pone.0005669. PMID: 19479078.

4. Boldsen JL. Epidemiological approach to the paleopathological diagnosis of leprosy. American Journal of Physical Anthropology. 2001;115(4):380-7.

5. Boldsen JL. Leprosy in Medieval Denmark—Osteological and epidemiological analyses. Anthropologischer Anzeiger. 2009:407-25.

6. Nerlich AG, Zink AR. Past Leprae. In: Raoult, D., Drancourt M. (eds) Paleomicrobiology. Past human infections. Springer-Verlag, Berlin. 2005:99 123.

7. Boldsen JL, Mollerup L. Outside St. Jørgen: leprosy in the medieval Danish city of Odense. American journal of physical anthropology. 2006; 130(3):344-51.

8. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009;41(12):1282-9. Epub 2009/11/03. doi: 10.1038/ng.477. PMID: 19881526.

9. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013;341(6142):179-83. Epub 2013/06/15. doi: 10.1126/science.1238286. PMID: 23765279.

10. Singh P, Benjak A, Carat S, Kai M, Busso P, Avanzi C, et al. Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2014;20(10):O619-22. Epub 2014/03/13. doi: 10.1111/1469-0691.12609. PMID: 24612452.

11. Das M, Chaitanya VS, Kanmani K, Rajan L, Ebenezer M. Genomic diversity in *Mycobacterium leprae* isolates from leprosy cases in South India. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2016;45:285-9. Epub 2016/10/23. doi: 10.1016/j.meegid.2016.09.014. PMID: 27642139.

12. Lavania M, Jadhav RS, Turankar RP, Chaitanya VS, Singh M, Sengupta U. Single nucleotide polymorphisms typing of *Mycobacterium leprae* reveals focal transmission of leprosy in high endemic regions of India. Clinical microbiology and infection :

the official publication of the European Society of Clinical Microbiology and Infectious Diseases. 2013;19(11):1058-62. Epub 2013/01/22. doi: 10.1111/1469-0691.12125. PMID: 23331718.

13. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the southern United States. The New England journal of medicine. 2011;364(17):1626-33. Epub 2011/04/29. doi: 10.1056/NEJMoa1010536. PMID: 21524213.

14. Sharma R, Singh P, Loughry WJ, Lockhart JM, Inman WB, Duthie MS, et al. Zoonotic Leprosy in the Southeastern United States. Emerg Infect Dis. 2015;21(12):2127-34. Epub 2015/11/20. doi: 10.3201/eid2112.150501. PubMed PMID: 26583204; PubMed Central PMCID: PMCPmc4672434.

15. Avanzi C, Del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744-7. Epub 2016/11/16. doi: 10.1126/science.aah3783. PMID: 27846605.

16. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(14):4459-64. Epub 2015/04/02. doi: 10.1073/pnas.1421504112. PMID: 25831531.

17. Mendum TA, Schuenemann VJ, Roffey S, Taylor GM, Wu H, Singh P, et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic. BMC genomics. 2014;15:270. Epub 2014/04/09. doi: 10.1186/1471-2164-15-270. PMID: 24708363.

18. Inskip SA, Taylor GM, Zakrzewski SR, Mays SA, Pike AW, Llewellyn G, et al. Osteological, biomolecular and geochemical examination of an early anglo-saxon case of lepromatous leprosy. PLoS One. 2015;10(5):e0124282. Epub 2015/05/15. doi: 10.1371/journal.pone.0124282. PMID: 25970602.

19. Mariotti V, Dutour O, Belcastro MG, Facchini F, Brasili P. Probable early presence of leprosy in Europe in a Celtic skeleton of the 4th–3rd century BC (Casalecchio di Reno, Bologna, Italy). International Journal of Osteoarchaeology. 2005;15(5):311-25.

20. Donoghue HD, Michael Taylor G, Marcsik A, Molnar E, Palfi G, Pap I, et al. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2015;31:250-6. Epub 2015/02/15. doi: 10.1016/j.meegid.2015.02.001. PMID: 25680828.

21. Belcastro M, Mariotti V, Facchini F, Dutour O. Leprosy in a skeleton from the 7th century necropolis of Vicenne-Campochiaro (Molise, Italy). International Journal of Osteoarchaeology. 2005;15(6):431-48.

22. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor protocols. 2010;2010(6):pdb.prot5448. Epub 2010/06/03. doi: 10.1101/pdb.prot5448. PMID: 20516186.

23. Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic acids research. 2012;40(1):e3. Epub 2011/10/25. doi: 10.1093/nar/gkr771. PMID: 22021376.

24. Herbig A, Maixner F, Bos KI, Zink A, Krause J, Huson DH. MALT: Fast alignment and analysis of metagenomic DNA sequence data applied to the Tyrolean Iceman. bioRxiv doi: http://dxdoiorg/101101/050559. 2016.

25. Maricic T, Whitten M, Paabo S. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One. 2010;5(11):e14004. Epub 2010/11/26. doi: 10.1371/journal.pone.0014004. PMID: 21103372.

26. Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prufer K, et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(37):14616-21. Epub 2007/08/24. doi: 10.1073/pnas.0704665104. PMID: 17715061.

27. Stoneking M, Krause J. Learning about human population history from ancient and modern genomes. Nature reviews Genetics. 2011;12(9):603-14. Epub 2011/08/19. doi: 10.1038/nrg3029. PMID: 21850041.

28. Sawyer S, Krause J, Guschanski K, Savolainen V, Paabo S. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One. 2012;7(3):e34131. Epub 2012/04/06. doi: 10.1371/journal.pone.0034131. PMID: 22479540.

29. Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, Paabo S. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic acids research. 2010;38(6):e87. Epub 2009/12/24. doi: 10.1093/nar/gkp1163. PMID: 20028723.

30. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nature genetics. 2007;39(12):1522-7.

31. Peltzer A, Jager G, Herbig A, Seitz A, Kniep C, Krause J, et al. EAGER: efficient ancient genome reconstruction. Genome biology. 2016;17:60. Epub 2016/04/03. doi: 10.1186/s13059-016-0918-z. PMID: 27036623.

32. Avanzi C, Busso P, Benjak A, Loiseau C, Fomba A, Doumbia G, et al. Transmission of Drug-Resistant Leprosy in Guinea-Conakry Detected Using Molecular Epidemiological Approaches. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2016;63(11):1482-4. Epub 2016/08/26. doi: 10.1093/cid/ciw572. PMID: 27558568.

33. Arentoft E. De spedalskes hospital: udgravninger af Sankt Jørgensgården i Odense: Odense Bys Museer; 1999.

34. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution. 2012;29:1969-73.

35. Kingman JFC. The coalescent. Stoch Proc Appl. 1982;13(3):235-48.

36. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Molecular biology and evolution. 2005;22(5):1185-92. Epub 2005/02/11. doi: 10.1093/molbev/msi103. PMID: 15703244.

37. Browne SG. The history of leprosy. In: Hastings RC, editor Leprosy 1st ed Edinburgh (Scotland) and New York: Churchill Livingstone. 1985:1–14.

38. Kohler K, Marcsik A, Zadori P, Biro G, Szeniczey T, Fabian S, et al. Possible cases of leprosy from the Late Copper Age (3780-3650 cal BC) in Hungary. PLoS One. 2017;12(10):e0185966. Epub 2017/10/13. doi: 10.1371/journal.pone.0185966. PMID: 29023477.

39. Boldsen JL. Leprosy and mortality in the Medieval Danish village of Tirup. American Journal of Physical Anthropology. 2005;126(2):159-68.

40. Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, et al. Molecular, ethno-spatial epidemiology of leprosy in China: novel insights for tracing leprosy in endemic and non endemic provinces. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases. 2013;14:361-8. Epub 2013/01/08. doi: 10.1016/j.meegid.2012.12.009. PMID: 23291419.

41. Yuan Y, Wen Y, You Y, Xing Y, Li H, Weng X, et al. Characterization of *Mycobacterium leprae* Genotypes in China--Identification of a New Polymorphism C251T in the 16S rRNA Gene. PLoS One. 2015;10(7):e0133268. Epub 2015/07/22. doi: 10.1371/journal.pone.0133268. PMID: 26196543.

42. Kim JP. SNP Genotypes of *Mycobacterium leprae* Isolated in Korea. Korean Lepr Bull. 2012;45:3-19.

43. Dabney J, Knapp M, Glocke I, Gansauge MT, Weihmann A, Nickel B, et al. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(39):15758-63. Epub 2013/09/11. doi: 10.1073/pnas.1314445110. PMID: 24019490.

44. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, et al. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. Nature protocols. 2009;4(6):960-74. Epub 2009/05/30. doi: 10.1038/nprot.2009.68. PMID: 19478811.

45. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95. Epub 2010/01/19. doi: 10.1093/bioinformatics/btp698. PMID: 20080505.

46. Okonechnikov K, Conesa A, Garcia-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32(2):292-4. Epub 2015/10/03. doi: 10.1093/bioinformatics/btv566. PMID: 26428292.

47. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20(9):1297-303.

48. Honore N, Perrani E, Telenti A, Grosset J, Cole ST. A simple and rapid technique for the detection of rifampin resistance in *Mycobacterium leprae*. International journal of leprosy and other mycobacterial diseases : official organ of the International Leprosy Association. 1993;61(4):600-4. Epub 1993/12/01 PMID: 8151192.

49. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame

genomic sequence comparison. Genome research. 2011;21(3):487-93. Epub 2011/01/07. doi: 10.1101/gr.113985.110. PMID: 21209072.

50. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution. 2016;33(7):1870-4. doi: 10.1093/molbev/msw054. PMID: 27004904.

Authors contributions:

VJS, CA, BKK, AN, KN, STC and JK designed the experiments. SI, DDP, GMT, PV, JL, AM, EM, GP, VM, AR, MGB, JLB, SM, HDD and SZ provided samples for analysis. VJS, CA, BKK, MB, ER, CU, GMT, PS and HDD performed the skeletal sampling. VJS, CA, BKK, MB, ER, and CU performed the ancient DNA experiments. AS, AH, AB and KN analyzed data. VJS, AB, CA, AS, AH, KN and JK wrote the manuscript with contributions from all co-authors. All authors read and approved the manuscript.

Acknowledgements:

This work is supported by the Fondation Raoul Follereau (S.TC.), the Swiss National Science Foundation grant IZRJZ3_164174 (S.T.C), the EMBO short-term fellowship (C.A), and the DFG Graduate School Human Development in Landscapes at Kiel University (B.K.-K., M.B., A.N.). Odense Bys Museer are thanked for granting access to the Odense St. Jørgen skeletons.

Supplementary tables:

Table S1: Details of the samples and screening results

Table S2: Details of the results of the genome-wide analysis

Table S3: Details of the SNPs per individual sample

Table S4: Lineage-specific protein-changing SNPs

Table S1: Details of the samples and screening results

Reference	average coverage (leprosy genes)	DMG 1st Base 5'	Screening PCR	metagenomic screening with MALT (#assigned reads to M. leprae/#raw reads)
Donoghue et al., 2015	224.58	0.1132	na	na
Donoghue et al., 2015	117.92	0.2049	na	na
Donoghue et al., 2015	69.47	0.1569	na	na
Mariotti et al., 2015	1.52	0.0339	na	na
	1.11	0.12	na	na
Donoghue et al., 2015	1.04	0	na	na
Donoghue et al., 2015	0.52	0.1176	na	na
Donoghue et al., 2015	2.77	0.0746	na	na
	0.39	0.1176	na	na
Inskip et al., 2015	305'602	0.1445	na	na
	700'063	0.1198	na	na
this study	na	na	positiv	2,624,974 / 19,464,646
this study	na	na	positiv	1,868,106 / 22,577,271
this study	na	na	positiv	424,828 / 28,125,125
this study	na	na	positiv	181,858 / 28,508,113
this study	na	na	positiv	52,858 / 21,352,057
this study	na	na	positiv	78,211 / 24,755,930

Table S2: Details of the results of the genome-wide analysis

Sample Name	# reads after C&M prior mapping	# mapped reads prior RMDup	# of Duplicates removed	Mapped Reads after RMDup	Endogenous DNA (%)	Mean Coverage	std. dev. Coverage	Coverage >= 1X in %	Coverage >= 5X in %	# SNPs	DMG 1st Base 3'	DMG 2nd Base 3'	DMG 1st Base 5'	DMG 2nd Base 5'	average fragment length	median fragment length	GC content in %
3077	5737174	1006722	482877	523845	17.547	10.006 6	8.2333	99.65	87.48	63	0.004	0.0026	0.0011	0.0009	62.43	59	56.74
Airak u3	58228079	14515514	9951478	4564036	24.929	151.84 55	44.6642	100	100	11 0	0.0036	0.0044	0.0032	0.0057	108.73	101	57.12
Body_ 188	68605185	22043205	17815330	4227875	32.130523 37	123.02 73	32.3442	100	100	10 1	0.0068	0.0054	0.0019	0.0008	95.1	86	56.92
Br492 3	1436010	1435321	3168	1432153	99.952	54.423 4	22.9519	99.89	97.81	11 7	0	0	0	0	124.19	150	57.8
Brw1 5-1	5115607	4849736	1702024	3147712	94.803	86.412 1	25.0615	100	99.97	12 7	0.0008	0.0006	0.0013	0.0008	89.72	101	57.63
Brw1 5-10	4353015	4217372	1325215	2892157	96.884	79.087 1	23.1094	100	99.96	12 6	0.001	0.0011	0.0007	0.0007	89.37	101	57.51
Brw1 5-20	2595709	2239722	434827	1804895	86.286	49.305 7	15.053	99.99	99.76	12 6	0.0005	0.0006	0.0007	0.0007	89.28	101	57.63
Brw1 5-25	3280926	3262150	882061	2380089	99.428	64.997 7	19.7714	100	99.89	12 5	0.0004	0.0006	0.0007	0.0005	89.25	101	57.38
Brw1 5-5	5886483	996349	133953	862396	16.926	23.661 1	8.281	99.94	98.84	12 4	0.0007	0.0006	0.0013	0.0009	89.67	101	57.39
GfC96	89513530	5012677	4226552	786125	5.6	19.795 9	24.1736	99.92	96.92	96	0.0042	0.0021	0.0031	0.0015	82.3	77	56.09
JK319 2	1178538	11580	8293	3287	0.9825733 24	0.0591	1.9306	0.45	0.15	45	0.0171	0.0274	0.0158	0.0224	58.78	54	57.6
JK319 3	27914905	64418	57044	7374	0.2307656 07	0.1318	3.1691	1.32	0.34	49	0.0209	0.0209	0.011	0.0155	58.4	50	58.8
JK319 4	1004184	9813	3327	6486	0.9772113 48	0.1245	1.617	7.19	0.16	10	0.0128	0.0099	0.0069	0.0081	62.72	58	56.35
JK319 5	14462110	99204	79101	20103	0.6859579 96	0.3905	2.9007	22	0.85	53	0.0102	0.0121	0.0072	0.0053	63.47	58	57.33
Jorgen_404	194327367	447991	50499	397492	0.231	10.136 9	3.8566	99.98	96.25	87	0.0028	0.001	0.0024	0.0006	83.35	83	57.86
Jorgen_427	28069326	180870	6342	174528	0.644	4.4234	2.3093	98.52	44.86	54	0.0036	0.0009	0.003	0.0006	82.83	82	57.66
Jorgen_472	1376434	1857	1	1856	0.135	0.0434	0.2184	4.16	0.01	0	0.0089	0.0019	0.0053	0	76.39	73	57.61
Jorgen_507	19203051	2594792	593647	2001145	13.512	55.059 2	10.0118	100	99.99	18 3	0.0022	0.001	0.0016	0.0004	89.92	93	57.46
Jorgen_533	27621188	412558	29513	383045	1.494	10.276 6	3.5191	99.97	96	10 8	0.0026	0.0012	0.0024	0.0006	87.68	91	57.2
Jorgen_722	1636080 99	486839	96612	390227	0.298	10.608 8	4.0328	99.99	97.54	11 8	0.0033	0.0016	0.0039	0.0009	88.85	83	57.38
Jorgen_749	22064466	1821831	454043	1367788	8.257	37.326 9	7.8586	100	99.99	91	0.0025	0.0011	0.002	0.0005	89.19	92	57.21
KH13 0047	37768028	11001	505	10496	0.029	0.2325	0.8823	19.52	0.06	12	0.0079	0.0087	0.0114	0.0092	72.4	68	58.65
KH13 0056	194197574	154300	29233	125067	0.079	3.087	3.3313	94.69	19.07	62	0.0031	0.0014	0.0027	0.0009	80.67	78	57.92
KH13 0060	80380766	44694	7342	37352	0.056	0.9767	1.8663	60.65	0.42	17	0.0034	0.0027	0.0027	0.0019	85.46	79	57.78
Kyoto 2	1436004	1435875	3314	1432561	99.991	54.448 7	22.9611	99.9	97.82	20 3	0	0	0	0	124.22	150	57.8

MI2-10	12972530	2953599	932571	2021028	22.768	55.037	15.1299	100	99.98	130	0.0056	0.0041	0.0009	0.0022	89	101	57.84
MI6-50	3093087	2851226	654377	2196849	92.181	59.8166	18.2112	100	99.85	131	0.0003	0.0005	0.0014	0.001	88.99	101	57.69
MI6-55	2985800	2606350	573581	2032769	87.292	55.4133	17.1575	100	99.82	129	0.0005	0.0006	0.0011	0.0008	89.09	101	57.55
ML_Thai53	1435926	1435925	2369	1433556	100	54.484	22.9272	99.89	97.83	42	0	0	0	0	124.21	150	57.8
NHDP-55	4101475	3769142	1148302	2620840	91.897	24.0576	7.9683	99.97	99.01	121	0.0009	0.0005	0.0003	0.0002	30	30	56.3
NHDP-98	6048880	5709066	2340675	3368391	94.382	35.8171	10.9711	100	99.77	116	0.0071	0.002	0.0006	0.0003	34.75	35	56.62
NHDP63	1436046	1436045	2444	1433601	100	54.486	22.9228	99.9	97.83	128	0	0	0	0	124.21	150	57.8
Refshale_16	43528407	12694753	7981097	4713656	29.164	119.8584	27.5433	100	100	83	0.0016	0.0013	0.0011	0.0007	83.1	76	57.42
S10	37991424	7242807	6305078	937729	19.064	17.0147	12.1812	100	99.71	NF	0.0083	0.0069	0.0048	0.0045	59.3	54	57.36
S11	47583911	30554537	24617589	5936948	64.212	163.2408	22.9881	100	100	NF	0.0178	0.0135	0.0011	0.0006	89.86	101	57.66
S13	9117933	1052599	381993	670606	11.544	15.1226	13.3934	99.96	97.65	NF	0.0148	0.0113	0.008	0.0059	73.7	74	57.37
S14	35058232	1899938	875282	1024656	5.419	19.483	16.4395	99.86	95.87	NF	0.0255	0.0481	0.0201	0.0446	62.16	58	56.96
S15	1846472	1004300	674274	330026	54.39	11.917	10.1757	99.88	95.31	226	0.0035	0.0024	0.0031	0.0022	118.01	114	57.87
S2	5577364	588428	173005	415423	10.55	10.2333	8.2062	99.77	88.76	NF	0.0126	0.0099	0.0035	0.0028	80.51	86	57.24
S9	3956035	1201468	832520	368948	30.371	15.0237	9.9031	99.93	97.65	168	0.0013	0.0016	0.0009	0.001	133.08	138	58.7
SK11	1679611	711636	244855	466781	42.36909618	8.2568	9.0318	98.68	73.02	45	0.0077	0.0044	0.0047	0.0008	57.81	55	55.63
SK14	4749524	1980390	433301	1547089	41.697	43.8653	19.3704	99.99	99.94	87	0.0014	0.0008	0.0012	0.0008	92.66	87	57.78
SK2	51741310	47872870	47244653	628217	92.523	12.511	12.2153	99.93	96.29	114	0.0054	0.0021	0.0056	0.0018	65.09	63	56.72
SK27	54407745	45501518	45305530	195988	83.63058973	3.8873	9.6435	81.05	22.85	6	0.0056	0.0028	0.0098	0.0037	64.82	62	54.5
SK8	10597941	1198268	304565	893703	11.307	21.0971	12.1367	99.99	99.25	87	0.002	0.0015	0.0016	0.001	77.15	76	56.87
T18	1611415	488693	145652	343041	30.32694868	7.2751	9.5459	98.01	64.2	25	0.0072	0.0057	0.0019	0.0008	69.31	66	56.05
TN	1435899	1435899	3091	1432808	100	54.4572	22.921	99.89	97.82	0	0	0	0	0	124.22	150	57.8

Table S3: Lineage-specific protein-changing SNPs

Lineage	Specific strain	Gene name	Modification	Function
Lineage 0	ancient strains	*ML0125*	482A>G, Tyr161Cys	Unknown
		ML0525	44G>T, Arg15Leu	Unknown
		subI	797C>T, Thr266Ile	Sulphate-binding lipoprotein
		era	97G>A, Val33Ile	GTPase
		sseA	538C>T, Leu180Phe	Thiosulfate sulfurtransferase
		mutB	1016C>T, Ala339Val	Methylmalonyl- CoA mutase
		ML2640c	580G>A, Ala194Thr	Unknown
		ctpI	22G>A, Asp8Asn	Cation transporter ATPase
	modern strains	*ML0411*	276C>A, His92Gln - 276C>T, His92His (Kyoto 2 and S10 have 276 C>A whereas S9 has 276 C>T)	Lysin rich antigen
Lineage 3	ancient strains	*none*	NA	NA
	modern strains	*gyrA*	3188C>T, Ser1063Phe	DNA gyrase subunit A
		asnB	971G>A, Arg324Gln	Asparagine synthetase
Lineage 4	ancient strains	*ML1334*	791G>C, Gly264Ala	Unknown
		ML0563	703C>G, Arg235Gly	Unknown
		ML0848	1677G>A, Met559Ile	ABC transporter
		lysX	86T>C, Val29Ala	lysyl-tRNA synthetase
	modern strains	*ML0411*	232G>A, Asp78Asn	Lysin rich antigen
		murX	763A>G, Ile255Val	Involved in cell wall formation

Chapter 4. Transmission and drug resistance

Chapter 4.1 – Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3

P. Singh[1], A. Benjak[1], S. Carat[2†], M. Kai[3], P. Busso[1], C. Avanzi[1], A. Paniz-Mondolfi[4,5], C. Peter[6], K. Harshman[6], J. Rougemont[2], M. Matsuoka[3] and S. T. Cole[1]

[1]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2]Bioinformatics and Biostatistics Core Facility, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[3]Leprosy Research Centre, National Institute of Infectious Diseases, Tokyo, Japan
[4]Departments of Biochemistry and Dermatopathology, Instituto de Biomedicina, Caracas, Venezuela
[5]Laboratory of Microbiology, Yale School of Medicine/Tuberculosis Reference Laboratory, VA Connecticut Healthcare, New Haven, CT, USA
[6]Centre for Integrative Genomics, University of Lausanne, Switzerland

2014. Clin Microbiol Infect. 20(10): 619-22

Contributions: DNA extraction and library preparation

doi.org/10.1111/1469-0691.12609

Genome-wide re-sequencing of multidrug-resistant *Mycobacterium leprae* Airaku-3

P. Singh[1], A. Benjak[1], S. Carat[2†], M. Kai[3], P. Busso[1],
C. Avanzi[1], A. Paniz-Mondolfi[4,5], C. Peter[6], K. Harshman[6],
J. Rougemont[2], M. Matsuoka[3] and S. T. Cole[1]

1) *Global Health Institute,* 2) *Bioinformatics and Biostatistics Core Facility, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland,*
3) *Leprosy Research Centre, National Institute of Infectious Diseases, Tokyo, Japan,* 4) *Departments of Biochemistry and Dermatopathology, Instituto de Biomedicina, Caracas, Venezuela,* 5) *Laboratory of Microbiology, Yale School of Medicine/Tuberculosis Reference Laboratory, VA Connecticut Healthcare, New Haven, CT, USA and* 6) *Centre for Integrative Genomics, University of Lausanne, Lausanne, Switzerland*

Abstract

Genotyping and molecular characterization of drug resistance mechanisms in *Mycobacterium leprae* enables disease transmission and drug resistance trends to be monitored. In the present study, we performed genome-wide analysis of Airaku-3, a multidrug-resistant strain with an unknown mechanism of resistance to rifampicin. We identified 12 unique non-synonymous single-nucleotide polymorphisms (SNPs) including two in the transporter-encoding *ctpC* and *ctpI* genes. In addition, two SNPs were found that improve the resolution of SNP-based genotyping, particularly for Venezuelan and South East Asian strains of *M. leprae*.

Corresponding author: S. T. Cole, Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland CH 1015 E-mail: stewart.cole@epfl.ch

Andrej Benjak and Solenne Carat contributed equally to this work.
†Current address: Service de Génétique Médicale, Laboratoire de Génétique moléculaire, 44093 Nantes, Cedex 01, France.

Keywords: Leprosy, molecular epidemiology, *Mycobacterium leprae*, rifampicin-resistance, single nucleotide polymorphism Original Submission: 29 January 2014; Accepted: 20 February 2014
Editor: D. Raoult
Article published online: 25 February 2014
Clin Microbiol Infect

Introduction

Despite a massive decline in leprosy prevalence in the last two decades, more than 200 000 new cases of leprosy are recorded each year globally, indicating active transmission of the infection [1]. Drug resistance to one or more anti-leprosy drugs has been reported but is rare. As the causative agent, *Mycobacterium leprae*, remains uncultivable, molecular drug susceptibility testing offers a practical alternative. This involves PCR-sequencing of the drug-resistance-determining regions of the *rpoB*, *folP1* and *gyrA* genes associated with resistance to rifampicin, dapsone and ofloxacin, respectively [2]. However, a previous study described low-levels of rifampicin resistance in *M. leprae* Airaku-3, isolated from a Japanese patient who relapsed after multidrug therapy [3]. This strain exhibited phenotypic resistance to rifampicin and dapsone in a mouse footpad assay. Although the dapsone-resistance was attributable to the known *folP1* mutation (Thr53Ile), Airaku-3 has a wild-type *rpoB* sequence; hence, an explanation for its rifampicin resistance is not available [2].

Alternative mechanisms of rifampicin resistance have been described in various bacterial species, for example, the *rox* gene mediated mono-oxygenation of rifampicin [4] and duplication of the *rpoB* gene as *rpoB2* in *Nocardia* [5]. There is no orthologue of *rox* in *Mycobacterium tuberculosis* and *M. leprae*. Over-expression of RNA polymerase-binding protein A causes low level rifampicin resistance in *Streptomyces coelicolor* [6]. Its *M. leprae* orthologue *ML1439* and other relevant genes *sigA* (*rpoT*) and *sigE* were all found to be wild-type in our PCR-based analysis. As per the *M. leprae* single nucleotide polymorphism (SNP) - genotyping scheme [7], Airaku-3 belongs to SNP subtype 1D, which is the predominant genotype in many countries like Bangladesh, India, Nepal, Madagascar, Malawi and the French West Indies [7– 9] and has a significant representation in Japan [7], Yemen, Venezuela [10] and the USA [11]. However, the genome of only one strain (S11-Inde2 from India) of this subtype has been sequenced [12]. Therefore, the Airaku-3 strain was selected for genome-wide sequencing using the Illumina platform to investigate the genetic basis for resistance to rifampicin and to improve the resolution of the existing SNP-genotyping scheme for *M. leprae*. In our present study, a bacillary suspension containing.10E+08 cells of *M. leprae* Airaku-3 [2] in 0.1 M NaOH was passed through a 1-mL insulin syringe (0.30-mm needle) three to five times, before DNA extraction [13]. Library preparation was by the TruSeq ChipSeq method (Illumina). Paired-end sequencing, with 101 cycles on a HiSeq2000 instrument provided a total of 174.2 million reads of which 25% could be aligned with the *M. leprae* TN

genome [14], yielding > times coverage, using BOWTIE [15]. *De novo* assembly of these reads was performed using VELVET [16] with *k*-mer size of 91. SNPs were identified as described previously [7,12] by comparison with the other *M. leprae* strains using PCR sequencing or genome comparisons [7,11,12] to identify the unique SNPs of the Airaku-3 genome (see Supporting information, Tables S1, S2). The phylogenetic trees were obtained with SPLITSTREE v.4.13.1 [17] using uncorrected *p* distances, the Neighbour Joining method and 1000 times bootstrapping (Fig. 1).

After *de novo* assembly of the reads the genome of *M. leprae* Airaku-3 comprised 114 contigs (*n*50 = 42 Kb) covering 98.85% of the TN genome. All the gaps corresponded to dispersed repeats. There was no evidence of structural variations compared with the *M. leprae* TN reference genome, thereby ruling out genome rearrangements or gene duplications as an explanation for rifampicin resistance in Airaku-3. Upon comparing the genome sequence of Airaku-3 with the *M. leprae* TN [14] reference genome, a total of 114 SNPs were identified. Of these, 46 SNPs were not present in any other indicated with a *. All bootstrap values are above 87 (average 99).

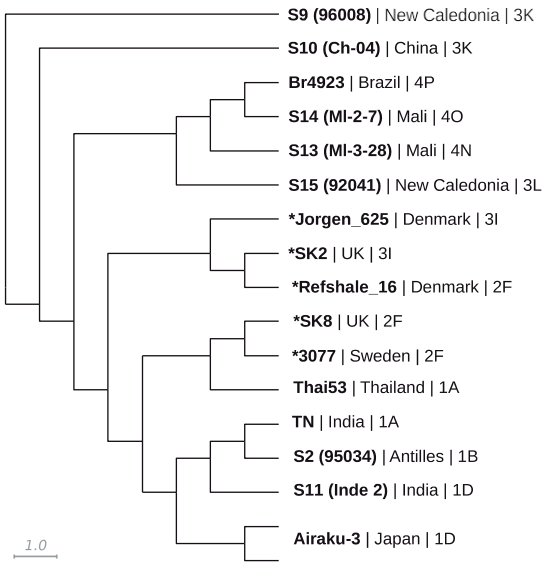


FIG. 1. Phylogenetic relationship of *Mycobacterium leprae* Airaku-3 with other *M. leprae* genomes [12]. For phylogeny, only the genomic positions in which all strains [12] had an unambiguous nucleotide call (674 positions) were considered. *Mycobacterium leprae* Airaku-3 is placed closest to another subtype 1D strain, S11-Inde2 from India, both of which share 12 SNPs that were absent in any other SNP genotypes. The geographic origins and the SNP genotypes are indicated against each strain. The ancient strains from Europe are indicated with a *. All bootstrap values are above 87 (average 99).

M. leprae genomes. These included two non-synonymous substitutions in genes with predicted transporter function belonging to the P-type ATPase family: 888973C>T in gene *ctpC* and 3209207G>A in *ctpI*. These two SNPs were also absent in the remaining *M. leprae* strains (Table S2).

Airaku-3 shared 12 SNPs with the other SNP-type 1D strain S11-Inde2 (Table S1). We analysed two of these SNPs (953582C>T and 3262657C>T) in 24 strains belonging to SNP subtype 1D from different countries (Table S2). SNP 3262657C>T correctly identified all of these 1D strains from the remaining 42 strains of other genotypes. Furthermore, the SNP 953582C>G distinguished the Venezuelan 1D strains (*n* = 10) from the rest of the 1D strains originating from eight different countries (Table S2). The 100 bp flanking region of this SNP in other mycobacterial outgroup species (*M. tuberculosis* complex, *M. avium* complex, *M. kansassi*, *M. marinum* and *M. ulcerans*) revealed a 'C' at the corresponding base, thereby defining the ancestral base. Therefore, Venezuelan 1D strains with a 'C' at position 953582 are designated as subtype 1D-1 while the remaining 1D strains with a derived base 'G' are termed 1D-2.

Though our present study does not reveal a clear explanation for the rifampicin resistance phenotype of Airaku-3, it has identified the unique SNPs, including two non-synonymous SNPs in transporter genes, *ctpC* and *ctpI*. However, a functional assay is required to determine whether either of these variant genes/transporters confers any degree of rifampicin resistance. Other transporters, for example, *drrA* (Rv2936), *pstB* (Rv0933) [18] and Rv1258c [19] reportedly confer low-levels of rifampicin resistance in *M. tuberculosis*. The Airaku-3 genome revealed no mutations in its *drrA* (ML2352c) whereas the orthologs of the remaining two genes (*pstB* and *ML1104c*) are pseudogenes in *M. leprae.* Rifampicin-resistant *M. tuberculosis* strains commonly possess compensatory mutations in the *rpoC* gene that restore their fitness [20]. This gene has a wild-type sequence in Airaku-3.

The comparative genomics of 1D strains in this study has discovered two useful markers: SNP3262657C>T defines the 1D genotype while the SNP953582C>G can further resolve them into 1D-1 and 1D-2. We have also identified a phylogeographic association of the 1D-1 genotype with Venezuela. These markers can be useful for molecular epidemiological studies in many countries where leprosy is endemic and the 1D genotype is predominant, and can also provide a possible explanation

for disease acquisition in other countries where sporadic cases are reported in the indigenous population or among immigrants or citizens who have lived in areas where leprosy is endemic. Previously, we had successfully resolved the subtype 3I strains into 3I-1 and 3I-2 using a similar approach [11]. Hence, our present study further exemplifies the value of genome-wide comparisons of a few strains to uncover reliable phylogeographic markers that can later be used for rapid PCR-based genotyping

Acknowledgements

We thank the patients and clinical staff from various leprosy clinics for their participation in this study and Marc Monot and Nadine Honoré for their initial contributions in developing the SNP-typing scheme for *M. leprae*. An academic license for $C_{ODON}C_{ODE}$ A_{LIGNER} software provided by Peter Richterich and Brian Campbell is also gratefully acknowledged. Part of the results of this work was presented at the 53rd Interscience Conference on Antimicrobial Agents and Chemotherapy (ICAAC) meeting at Denver (6–9 September 2013, Poster presentation number C1-512). Illumina reads of *M. leprae* Airaku-3 can be accessed at the National Center for Biotechnology Information Sequence Read Archive under accession no. SRR113282.

Funding

We are grateful to the Fondation Raoul Follereau, the Brazilian–Swiss Joint Research Project (Swiss National Science Foundation) and the Global Health Institute, EPFL, Lausanne, Switzerland for financial support.

Contribution to Authorship

PS and STC designed the study with the help of MK and MM; PS, PB, CP, CA and AP-M did the sample preparation, sequencing and analysis with guidance from KH and STC; AB, SC, JR and PS performed the bioinformatic analysis; PS wrote the manuscript with input from AB, STC and other co-authors. All authors read and approved the final version of the manuscript.

Transparency Declaration

The authors declare no conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. List of all single nucleotide polymorphisms (SNPs) in *Mycobacterium leprae* Airaku-3 genome compared to the TN genome and their distribution in previously described genomes [7,12]. The unique SNPs and their effect in the Airaku-3 genome are shown in bold, with the SNPs in transporter proteins shown in blue highlight.

Table S2. *Mycobacterium leprae* isolates of various genotypes from different geographic origins, which were tested for the subtype 1D specific single nucleotide polymorphisms. The combination of 'T and C' at positions 3262657 and 953582, respectively, is characteristic of the Venezuelan 1D-1 genotype whereas a combination of 'T and G' represents the 1D-2 genotype.

References

1. World Health Organization Global leprosy: update on the 2012 situation. *Wkly Epidemiol Rec* 2013; 88: 365–380.
2. Maeda S, Matsuoka M, Nakata N, et al. Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrob Agents Chemother* 2001; 45: 3635–3639.
3. Matsuoka M. History and characteristics of isolates maintained at the leprosy research center. *Nihon Hansenbyo Gakkai Zasshi* 2010; 79: 247– 256.
4. Hoshino Y, Fujii S, Shinonaga H, et al. Monooxygenation of rifampicin catalyzed by the rox gene product of *Nocardia farcinica*: structure elucidation, gene identification and role in drug resistance. *J Antibiot (Tokyo)* 2010; 63: 23–28.
5. Ishikawa J, Chiba K, Kurita H, Satoh H. Contribution of rpoB2 RNA polymerase beta subunit gene to rifampin resistance in *Nocardia* species. *Antimicrob Agents Chemother* 2006; 50: 1342–1346.
6. Newell KV, Thomas DP, Brekasis D, Paget MSB. The RNA polymerase-binding protein RbpA confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. *Mol Microbiol* 2006; 60: 687–696.
7. Monot M, Honore N, Garnier T, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat Genet* 2009; 41: 1282–1289.
8. Kuruwa S, Vissa V, Mistry N. A study of the distribution of *Mycobacterium leprae* strains among cases in a rural and urban population of Maharashtra, India. *J Clin Microbiol* 2012; 50: 1406– 1411.
9. Lavania M, Jadhav RS, Turankar RP, Chaitanya VS, Singh M, Sengupta U. Single nucleotide polymorphisms typing of *Mycobacterium leprae* reveals focal transmission of leprosy in high endemic regions of India. *Clin Microbiol Infect* 2013; 19: 1058–1062.
10. Singh P, Busso P, Paniz-Mondolfi A, et al. Molecular drug susceptibility testing and genotyping of *Mycobacterium*

leprae from South America. *Antimicrob Agents Chemother* 2011; 55: 2971–2973.
11. Truman RW, Singh P, Sharma R, et al. Probable zoonotic leprosy in the southern United States. N Engl J Med 2011; 364: 1626–1633.
12. Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science 2013; 341: 179–183

Table S1: List of all single nucleotide polymorphisms (SNPs) in *Mycobacterium lepraeAiraku-3* genome compared to the TN genome and their distribution in previously described genomes [7,12]. The unique SNPs and their effect in the Airaku-3 genome are shown in bold, with the SNPs in transporter proteins shown in blue highlight.

TN genomic coordinate	TN base	Airaku-3 base	ORF	Gene name	AA change	Previous reference of the SNP	Description
5066	G	T				Not described previously	Intergenic
206721	G	A	ML0145	ML0145c		Not described previously	Possible transposase remnant (pseudogene)
361545	G	C	ML0279	ML0279c	P>A	Not described previously	Conserved hypothetical protein
369492	C	T	ML0288	ML0288		Not described previously	Possible transposase remnant (pseudogene)
417000	T	G	ML0327	esxS		Not described previously	Conserved hypothetical protein EsxS (pseudogene)
430198	C	T	ML0340	fadE34		Not described previously	Probable acyl-CoA dehydrogenase FadE34 (pseudogene)
477016	G	A	ML0383	ML0383	G>R	Not described previously	Conserved hypothetical protein
506336	G	A	ML0408	ML0408		Not described previously	Possible IS1081 transposase (pseudogene)
509065	G	A	ML0411	ML0411	G>D	Not described previously	Serine-rich antigen
593138	C	T	ML0488	secF	Syn	Not described previously	Probable protein-export membrane protein SecF
638512	C	T	ML0526	ML0526c		Not described previously	Probable oxidoreductase (pseudogene)
728810	G	A				Not described previously	Intergenic
878760	C	T	ML0737	fadE25	A>V	Not described previously	Probable acyl-CoA dehydrogenase FadE25
888972	C	T	ML0747	ctpC	G>R	Not described previously	Probable metal cation-transporting P-type ATPase C CtpC
924598	A	G				Not described previously	Intergenic
1069840	C	T	ML0903	ML0903c	R>Q	Not described previously	conserved hypothetical protein
1120609	C	T	ML0952	ML0952		Not described previously	Conserved hypothetical protein (pseudogene)
1204421	C	T	ML1039	ML1039		Not described previously	Possible oxidoreductase (pseudogene)
1220708	T	C				Not described previously	Intergenic
1225568	G	T	ML1060	ML1060c		Not described previously	Conserved hypothetical protein (pseudogene)
1442144	C	T	ML1219	ML1219		Not described previously	Conserved hypothetical protein (pseudogene)
1475671	G	A	ML1240; ML1239			Not described previously	Unknown (pseudogene)
1481100	C	T				Not described previously	Intergenic
1619055	G	T	ML1359	ppnK	Syn	Not described previously	Inorganic polyphosphate/ATP-NAD kinase PpnK (Poly(P)/ATP NAD kinase)
1704410	G	A	ML1419	ML1419c	L>F	Not described previously	regulatory protein
1723598	G	A				Not described previously	Intergenic
1792555	C	A				Not described previously	Intergenic
1920611	G	A	ML1595	ML1595		Not described previously	Probable transcriptional regulatory protein (pseudogene)
1933895	G	A				Not described previously	Intergenic
1977878	G	A	ML1640	cobC		Not described previously	Possible aminotransferase CobC (pseudogene)
2020721	G	A	ML1676	thiL	P>S	Not described previously	Probable thiamine-monophosphate kinase thiL (thiamine-phosphate kinase)
2355529	G	A	ML1971	ML1971c		Not described previously	Possible transposase remnant (pseudogene)
2395657	C	G	ML2002	fadB5		Not described previously	Possible oxidoreductase FadB5 (pseudogene)
2408447	C	G	ML2015	cinA	G>A	Not described previously	PROBABLE CINA-LIKE PROTEIN CINA
2717888	C	A	ML2293	dppA		Not described previously	Probable periplasmic dipeptide-binding lipoprotein DppA (pseudogene)
2736012	C	T	ML2310	ML2310		Not described previously	Probable lyase, cysteine metabolism (pseudogene)
2740755	A	G				Not described previously	Intergenic
2878040	C	A	ML2406	menA	Q>K	Not described previously	1,4-DIHYDROXY-2-NAPHTHOATE OCTAPRENYLTRANSFERASE MENA (DHNA-OCTAPRENYLTRANSFERASE)
2977226	C	T	ML2499	ML2499		Not described previously	Probable carbon monoxide dehydrogenase (pseudogene)
2977229	A	T	ML2499	ML2499		Not described previously	Probable carbon monoxide dehydrogenase (pseudogene)
3057161	C	T				Not described previously	Intergenic
3096072	C	G	ML2591	mce1C	A>G	Not described previously	MCE-FAMILY PROTEIN MCE1C
3109982	A	G	ML2606	bglS		Not described previously	Probable beta-glucosidase BglS (pseudogene)
3113891	G	A	ML2608	ilvD	Syn	Not described previously	PROBABLE DIHYDROXY-ACID DEHYDRATASE ILVD (DAD)
3201250	G	A				Not described previously	Intergenic
3209198	G	A	ML2671	ctpI	A>T	Not described previously	PROBABLE CATION-TRANSPORTER ATPASE I CTPI
492614	C	A	ML0395	guaA	Syn	Shared only with another 1D strain, Inde2 (Ref 12)	Probable GMP synthase [glutamine-hydrolyzing] GuaA (Glutamine amidotransferase) (GMP synthetase)
639580	A	G				Shared only with another 1D strain, Inde2 (Ref 12)	Intergenic
953582	C	G	ML0803	ML0803	R>G	Shared only with another 1D strain, Inde2 (Ref 12)	Probable two-component sensor kinase
1343119	T	C	MLP000017		M>T	Shared only with another 1D strain, Inde2 (Ref 12)	23S ribosomal RNA
2166647	C	G				Shared only with another 1D strain, Inde2 (Ref 12)	Intergenic
2181587	G	A	ML1800	mutA	A>V/start	Shared only with another 1D strain, Inde2 (Ref 12)	PROBABLE METHYLMALONYL-CoA MUTASE SMALL SUBUNIT MUTA (MCM)
2181893	G	A	ML1800	mutA	S>L/start	Shared only with another 1D strain, Inde2 (Ref 12)	PROBABLE METHYLMALONYL-CoA MUTASE SMALL SUBUNIT MUTA (MCM)
2468622	C	T	ML2075	ML2075c	G>R	Shared only with another 1D strain, Inde2 (Ref 12)	conserved hypothetical protein
3016895	C	A	ML2535	eccC3	V>L	Shared only with another 1D strain, Inde2 (Ref 12)	ESX CONSERVED COMPONENT ECCC3. POSSIBLE MEMBRANE PROTEIN.
3070616	A	C	ML2572	ML2572		Shared only with another 1D strain, Inde2 (Ref 12)	Probable conserved transmembrane protein (pseudogene)
3262657	C	T	ML2706	parB	Syn	Shared only with another 1D strain, Inde2 (Ref 12)	PROBABLE CHROMOSOME PARTITIONING PROTEIN PARB
3267618	G	A	ML2713	rpmH	Syn	Shared only with another 1D strain, Inde2 (Ref 12)	50S RIBOSOMAL PROTEIN L34 RPMH
8453	T	C	ML0006	gyrA	L>P	Described in Ref 12	Probable DNA gyrase (subunit A) GyrA (DNA topoisomerase (ATP-hydrolysing)) (DNA topoisomerase II) (Type II DNA topoisomerase)
52851	T	C	ML0042	eccE1	W>R	Described in Ref 12	ESX CONSERVED COMPONENT ECCE1. POSSIBLE MEMBRANE PROTEIN.
61425	A	G	ML0049	esxA	M >T	Described in Ref 12	Probable 6 kDa early secretory antigenic target homolog EsxA (ESAT-6-like protein) (L-ESAT)
61530	C	T	ML0049	esxA	R>Q	Described in Ref 12	Probable 6 kDa early secretory antigenic target homolog EsxA (ESAT-6-like protein) (L-ESAT)
77864	T	C	ML0061	gltB	Syn	Described in Ref 12	Probable ferredoxin-dependent glutamate synthase (large subunit) GltB (L-glutamate synthase) (L-glutamate synthetase) (NADH-glutamate synthase) (NADPH-GOGAT)
264501	A	C	ML0197	ML0197c		Described in Ref 12	Probable helicase (pseudogene)
296853	C	T	ML0224	folP1	T>I	Described in Ref 12	Probable dihydropteroate synthase 1 FolP1 (DHPS 1) (dihydropteroate pyrophosphorylase 1) (dihydropteroate dipyrophosphorylase 1)
298572	A	G	ML0227	ML0227	Syn	Described in Ref 12	Probable conserved secreted or membrane protein
310911	G	A	ML0237	ML0237		Described in Ref 12	Hypothetical protein (pseudogene)
313361	A	G	ML0238	metS	L>P	Described in Ref 12	Probable methionyl-tRNA synthase MetS
347385	T	C	ML0266	ML0266		Described in Ref 12	Conserved hypothetical protein (pseudogene)
383599	C	G	ML0301	mutT3		Described in Ref 12	Possible mutator protein MutT3 (pseudogene)
481476	A	G	ML0387	guaB2	Syn	Described in Ref 12	Probable inosine-5'-monophosphate dehydrogenase GuaB2 (IMP dehydrogenase) (IMPDH) (IMPD)
485138	T	C	ML0389	choD	L>P	Described in Ref 12	Probable cholesterol oxidase precursor ChoD (cholesterol-O2 oxidoreductase)
494674	T	G	ML0397	ML0397c	E>D	Described in Ref 12	Possible transporter protein
508481	T	C	ML0410	PE3	V>A	Described in Ref 12	PE-family protein
509325	C	G	ML0411	ML0411	H>D	Described in Ref 12	Serine-rich antigen
533403	A	G	ML0433	fadD35		Described in Ref 12	Probable fatty-acid-CoA synthase FadD35 (pseudogene)
657736	G	C	ML0541	gmk	Syn	Described in Ref 12	Probable guanylate kinase Gmk (GMP kinase).
694090	T	C	ML0569	ML0569c	T>A	Described in Ref 12	Possible conserved exported protein
840584	A	G	ML0702	ML0702c		Described in Ref 12	Probable membrane protein (pseudogene)
890453	A	G	ML0747	ctpC	V>A	Described in Ref 12	Probable metal cation-transporting P-type ATPase C CtpC
904824	G	C	ML0763	pmmA	Syn	Described in Ref 12	probable phosphomannomutase PmmA (PMM) (PHOSPHOMANNOSE MUTASE)
1087397	T	C	ML0917	ftsZ	Syn	Described in Ref 12	cell division protein FtsZ
1143423	T	C	ML0975	ML0975c		Described in Ref 12	Probable dehydrogenase (pseudogene)
1155582	T	G	ML0988	recX	Syn	Described in Ref 12	PROBABLE REGULATORY PROTEIN RECX
1227631	C	A	ML1062	fadD6		Described in Ref 12	Probable fatty-acid-CoA ligase ligase FadD6 (pseudogene)
1257185	T	C				Described in Ref 12	Intergenic
1265267	T	G	ML1097	ML1097		Described in Ref 12	Probable drug-transport integral membrane protein (pseudogene)
1324009	C	G	ML1132	rho	Q>E	Described in Ref 12	PROBABLE TRANSCRIPTION TERMINATION FACTOR RHO HOMOLOG
1348426	T	C				Described in Ref 12	Intergenic
1351149	C	G				Described in Ref 12	Intergenic
1407885	A	G				Described in Ref 12	Intergenic
1529088	A	G	ML1284	cydA		Described in Ref 12	Probable integral membrane cytochrome D ubiquinol oxidase (subunit I) CydA (pseudogene)
1553354	T	G	ML1304	ansP1		Described in Ref 12	POSSIBLE L-ASPARAGINE PERMEASE ANSP1 (L-ASPARAGINE TRANSPORT PROTEIN)
1579109	T	G	ML1327	ML1327c		Described in Ref 12	Conserved hypothetical protein (pseudogene)
1614069	T	C	ML1353	ML1353c		Described in Ref 12	Conserved hypothetical protein (pseudogene)
1625045	T	G	ML1363	pyrG	Syn	Described in Ref 12	Probable CTP synthase PyrG
1643162	T	C	ML1378	ML1378		Described in Ref 12	conserved hypothetical protein (pseudogene)
1677493	G	C	ML1397	tsnR	Syn	Described in Ref 12	Possible 23S rRNA methyltransferase TsnR
1701590	G	C				Described in Ref 12	Intergenic
1843283	C	A	ML1528	ML1528		Described in Ref 12	Conserved hypothetical protein (pseudogene)
2043287	A	G	ML1694	ilvC	Syn	Described in Ref 12	PROBABLE KETOL-ACID REDUCTOISOMERASE ILVC (Acetohydroxy-acid isomeroreductase) (Alpha-keto-beta-hydroxylacil reductoisomerase)
2100523	C	G	ML1740	ML1740c	A>P	Described in Ref 12	PROBABLE SHORT CHAIN ALCOHOL DEHYDROGENASE/REDUCTASE
2104127	T	C				Described in Ref 12	Intergenic
2155013	T	G	ML1778	ML1778c		Described in Ref 12	Possible transcriptional regulatory protein (pseudogene)
2174865	G	C				Described in Ref 12	Intergenic
2209759	A	G	ML1824	ML1824c		Described in Ref 12	Putative group II intron maturase-related protein (pseudogene)
2553176	T	G	ML2149	moaA2		Described in Ref 12	Probable molybdenum cofactor biosynthesis protein A MoaA2 (pseudogene)
2706236	T	G	ML2281	ML2281c		Described in Ref 12	Probable monooxygenase (pseudogene)
2818521	T	C	ML2357	ppsA		Described in Ref 12	PHENOLPTHIOCEROL SYNTHESIS TYPE-I POLYKETIDE SYNTHASE PPSA
2964999	T	C	ML2490	clpB	Syn	Described in Ref 12	PROBABLE ENDOPEPTIDASE ATP BINDING PROTEIN (CHAIN B) CLPB (CLPB PROTEIN) (HEAT SHOCK PROTEIN F84.1)
2990150	T	C				Described in Ref 12	Intergenic
3016175	T	C	ML2534	PE10	Syn	Described in Ref 12	PE-family protein
3175296	A	C				Described in Ref 12	Intergenic
1890243	C	G				Described in Ref 12	Intergenic

Table S2: *Mycobacterium leprae* isolates of various genotypes from different geographic origins, which were tested for the subtype 1D specific single nucleotide polymorphisms. The combination of 'T and C' at positions 3262657 and 953582, respectively, is characteristic of the Venezuelan 1D-1 genotype whereas a combination of 'T and G' represents the 1D-2 genotype.

Number	Strain	SNP Subtype	Geographic origin	SNP 3262657 C>T	SNP 953582 C>G	Ref/ Source
1	TN	1A	India	C	C	[14]
2	Thai53	1A	Thailand	C	C	[7]
3	S2-95034	1B	Antilles	C	C	[12]
4	Refshale_16	2F	Denmark*	C	C	[12]
5	3077	2F	Sweden*	C	C	[12]
6	SK8	2F	UK*	C	C	[12]
7	85054	3I	Antilles	C	C	This study
8	BP	3I	Brazil	C	C	This study
9	BrMM4	3I	Brazil	C	C	This study
10	BrMM5	3I	Brazil	C	C	This study
11	BrPS5	3I	Brazil	C	C	This study
12	Jorgen_625	3I	Denmark*	C	C	[10]
13	NHDP55	3I	USA	C	C	[11]
14	NHDP63	3I	USA	C	C	[7]
15	NHDP98	3I	Mexico	C	C	[11]
16	SK2	3I	UK*	C	C	[12]
17	NHDP63	3I	USA	C	C	[7]
18	WA [I-30]	3I	USA	C	C	[11]
19	S10-Ch-04	3K	China	C	T	[12]
20	S9-96008	3K	New Caledonia	C	C	[12]
21	S15-92041	3L	New Caledonia	C	C	[12]
22	97016	4N	Antilles	C	C	This study
23	Bn3-11	4N	Benin	C	C	This study
24	Br1	4N	Brazil	C	C	This study
25	JPN	4N	Brazil	C	C	This study
26	Africa	4N	Ethiopia	C	C	This study
27	ML2-5	4N	Mali	C	C	This study
28	ML4-34	4N	Mali	C	C	This study
29	S13-Ml-3-28	4N	Mali	C	C	[12]
30	Mex7R	4N	Mexico	C	C	This study
31	Mx78	4N	Mexico	C	C	This study
32	Ng1	4N	Niger	C	C	This study
33	V3-1	4N	Venezuela	C	C	This study
34	Mx7Vera	4N/O	Mexico	C	C	This study
35	97011	4O	Antilles	C	C	This study
36	S14-Ml-2-9	4O	Mali	C	C	[12]
37	Br4923	4P	Brazil	C	C	[7]
38	BrMM1	4P	Brazil	C	C	This study
39	BrMM2	4P	Brazil	C	C	This study
40	BrSBSn	4P	Brazil	C	C	This study

Number	Strain	SNP Subtype	Geographic origin	SNP 3262657 C>T	SNP 953582 C>G	Ref/ Source
41	Fio3	4P	Brazil	C	C	This study
42	Ger1	3I	Brazil[Ger]	C	C	This study
43	V1-84	1D-1	Venezuela	T	C	This study
44	V1-86	1D-1	Venezuela	T	C	This study
45	V108	1D-1	Venezuela	T	C	This study
46	V207	1D-1	Venezuela	T	C	This study
47	V209	1D-1	Venezuela	T	C	This study
48	V211	1D-1	Venezuela	T	C	This study
49	V213	1D-1	Venezuela	T	C	This study
50	V32	1D-1	Venezuela	T	C	This study
51	VB11	1D-1	Venezuela	T	C	This study
52	VB80	1D-1	Venezuela	T	C	This study
53	Bang-112	1D-2	Bangladesh	T	G	This study
54	S11-Inde2	1D-2	India	T	G	[12]
55	India-823	1D-2	India	T	G	This study
56	MFP5	1D-2	India	T	G	This study
57	Airaku-3	1D-2	Japan	T	G	This study
58	Japan-27	1D-2	Japan	T	G	This study
59	Japan-29	1D-2	Japan	T	G	This study
60	Japan-30	1D-2	Japan	T	G	This study
61	Japan-31	1D-2	Japan	T	G	This study
62	B14853	1D-2	Nepal	T	G	This study
63	Mdg-B107	1D-2	Madagascar	T	G	This study
64	2936	1D-2	Malawi	T	G	This study
65	Pak1	1D-2	Pakistan	T	G	This study
66	Ye1-2	1D-2	Yemen	T	G	This study

* ancient *M. leprae* strains; [Ger] a German national who resided in Brazil for a long duration.

Chapter 4.2 – Transmission of drug-resistant leprosy in Guinea-Conakry detected using molecular epidemiological approaches

Charlotte Avanzi,[1] Philippe Busso,[1] Andrej Benjak,[1] Chloé Loiseau,[1] Abdoulaye Fomba,[2] Glodia Doumbia,[2] Idrissa Camara,[3] André Lamou,[3] Gouressy Sock,[3] Tiguidanké Drame,[3] Mamadou Kodio,[2] Fatoumata Sakho,[3] Samba O. Sow,[2] Stewart T. Cole,[1#] and Roch Christian Johnson[4#]

[1]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Switzerland
[2]Centre national d'Appui à la lutte contre la Maladie, Bamako, Mali
[3]Programme National Lèpre de Guinée, Conakry, Guinea
[4]Centre InterFacultaire de Formation et de Recherche en Environnement pour le Développement Durable University of Abomey-Calavi, Cotonou, Benin

[#] Corresponding author

2016. Clin Infect Dis. 63(11): 1482-1484

Contributions: DNA extraction, library preparation, interpretation of results, manuscript preparation and submission

BRIEF REPORT

Abstract: Molecular drug susceptibility testing was performed on skin biopsies from 24 leprosy patients from Guinea-Conakry for the first time. We identified primary drug resistance in 4 cases and a dapsone-resistant cluster caused by the same strain. Primary transmission of drug-resistant *Mycobacterium leprae*, including a rifampicin-resistant strain, is reported.

Keywords. drug resistance; leprosy; Guinea-Conakry.

Despite a remarkable decline in the prevalence of leprosy following the World Health Organization's (WHO) global implementation of multidrug therapy (MDT) in the 1980s [1, 2], well over 200 000 new leprosy cases are recorded annually worldwide. The incidence rate has stagnated since 2005, indicating continuation of active transmission of the disease. First-line drugs used for MDT against leprosy include dapsone, rifampicin, and clofazimine; second-line drugs include ofloxacin, minocycline, and clarithromycin. Because of the lack of effective alternative antileprosy drugs, resistance to the first-line drugs could seriously affect leprosy-control programs. The resistance of *Mycobacterium leprae* to antileprosy drugs has been observed in several leprosy-endemic regions, which is reported by the Global Sentinel Surveillance for Drug Resistance in Leprosy program coordinated by WHO [3].

Drug-resistant leprosy can occur either by transmission of a resistant strain (primary resistance) or by mutation of the wild-type drug-susceptible strain during therapy (secondary resistance). Only a few cases of primary resistance to dapsone have been reported to date [4]. However, primary rifampicinresistant cases are more often described, and they are a cause of concern due to the limited availability of second-line drugs for leprosy [4].

Leprosy is endemic in Guinea-Conakry; 313 cases were reported in 2014, making it a country with a moderate leprosy burden. No information about drug resistance in this region is available. Our aim was to obtain a preliminary estimation of the drug-resistance levels between relapses and primary leprosy cases from Guinea-Conakry using molecular methods.

DNA was extracted from ethanol-fixed skin biopsy specimens using the "freeze-boiling" method. The presence of mutations in drug resistance–determining regions (DRDRs) of *rpoB, folP1*, and *gyrA* genes, associated with rifampicin, dapsone, and quinolone resistance, respectively, was tested using *M. leprae* DRDR primers and a polymerase chain reaction (PCR) sequencing method [5]. Sequences obtained from the ABI3130xl genetic analyzer (Thermo Fisher Scientific, Waltham, Massachusetts) were aligned onto the *M. leprae* Tamil Nadu (TN) reference strain sequence using CodonCode Aligner software (Dedham, Massachusetts) in order to identify mutations.

We performed whole-genome sequencing of 3 samples. For this, an additional biopsy was first digested using collagenase/ dispase (Roche, Switzerland) and trypsin (AppliChem, Germany); DNA was extracted with the QIAmp microbiome kit (Qiagen, Netherland) using an adapted method. Briefly, after human cell lysis and DNase treatment, bacilli were lysed with a mixture of 20 mg/mL of proteinase K (Qiagen, Netherland) and lysozyme (Sigma, St Louis, Missouri) prior to DNA purification on silica-based columns. DNA samples were subsequently sonicated to 400 bp-long fragments using S220 Covaris instruments. Illumina libraries were prepared using the Kapa hyper prep kit (Kapa Biosystems, Wilmington, Massachusetts) with the PentAdpters (Pentabases, Denmark) indexed adapters. After quality control (library concentration and fragment analyzer),

libraries were multiplexed and sequenced as 100 base-long single-end reads on an Illumina HiSeq 2500 instrument. Raw reads were adapter trimmed and quality trimmed with Trimmomatic v0.33 [6] and mapped onto the *M. leprae* TN reference genome (NCBI a.n. AL450380.1) with Bowtie2 [7] (version 2.2.5) followed by variant calling using VarScan v2,3,9 [8]. Single nucleotide polymorphisms (SNPs) were called after applying quality checks, namely, a minimum overall coverage of 5 nonduplicated reads, a minimum of 3 nonduplicated reads supporting the SNP, a mapping quality score >8, a base quality score >15, and a SNP frequency above 80%. These cutoffs were chosen to avoid false positives. For phylogenetic analysis, the SNP type and subtype were inferred using the SNP-typing system described by Monot et al as well as the variable number of tandem repeats (VNTRs) profile [9, 10]. Finally, comparative genomics was performed on the 3 strains to identify unique and shared variants. Phylogenetic analysis was done as described by Schuenemann et al [11]. Illumina reads of *M. leprae* can be accessed from the Sequence Read Archive database (National Center for Biotechnology Information) under accession numbers SRX1677242 (Ml2-10), SRX1677243 (Ml6-50), and SRX1677244 (Ml6-55).

A total of 24 samples were collected between 2012 and 2015, including 1 from a relapse case with no history of monotherapy (Table 1) and 23 from new cases. Patients originated from regions of Boke (n = 4), Faranah (n = 2), Nzérékaré (n = 1), Mamou (n= 1), Kankan (n= 3), and Kindia (n = 13). There were 16 males and 8 females. All patients were multibacillary (MB) cases including 6 with grade 2 disabilities. The bacillary index ranged from 1+ to 6+ but was not available for 6 patients. MDT was prescribed for 12 months according to the WHO recommendation for MB leprosy.

PCR sequencing of DRDRs revealed a mutation in *rpoB* codon 456 (Ser(TCG) → Leu(TTG)) in 1 primary case. The patient was a male from Boke who presented with anesthetic patches on the back and trunk. A history of leprosy within the family was not available. This mutation has been associated with high-level rifampicin-resistant leprosy [3]. In addition, a missense mutation at codon 53 in *folP1* (Thr(ACC) → Arg (AGA)) associated with dapsone-resistant leprosy was observed in 3 primary cases [3]. This dinucleotide mutation has been reported in clinical specimens from India and the Philippines (relapse case) as conferring high or intermediate resistance levels [12, 13]. This mutation has not been identified elsewhere. No mutations were identified in the *gyrA* DRDR associated with ofloxacin resistance (Table 1).

We investigated the 3 cases with the same *folP1* mutation in more detail. The 3 patients shared the same prefecture of residence, Mandiana, in the Kankan region in the eastern part of the country. They all had MB leprosy and were diagnosed in 2011 (patient Ml2-10) and 2013 (patients Ml6-50 and Ml6-55). In addition, Ml2-10 and Ml6-55 were siblings; however, we could not establish contact with case Ml6-50.

Table 1. Results of the Drug-Resistance Surveillance in the GuineaConakry

Drug Resistance	New Cases	Relapse Cases
Mutation in *rpoB* (Ser456Leu)	1	0
Mutation in *folP1* (Thr53Arg)	3	0
Mutation in *gyrA*	0	0
Drug susceptible	19	1
Total	23	1

We obtained sufficient whole genome read coverage (54-58X) for the 3 strains for downstream analyses. All 3 strains belonged to SNP type 4N,which is a common genotype in West Africa and Brazil [9]. Phylogeny based on whole genome SNP alignments

unambiguously placed the 3 strains in 1 individual subbranch compared with the previously published *M. leprae* genomes [14]. We identified only 2 SNPs between the 3 genomes: 1 SNP was found in Ml2-10 but not in Ml6-50 and Ml6-55 and 1 SNP was found in Ml2-10 and Ml6-55 but not in Ml6-50. In addition, we analyzed 31 microsatellite and 11 minisatellite loci in the 3 resistant strains and compared them with the VNTR profiles of S13, Thai-53, Airaku-3, and Zensho-4 strains, respectively, genotypes 4N, 1A, 1D, and 3K [10, 11]. The 3 strains displayed the same VNTR profile (Supplementary Tables 1 and 2) except for 4 microsatellites loci $(G)^{11}$, $(G)^{12}$, $(C)^{20}$, and $(AGA)^{20}$ (Supplementary Table 1), thus corroborating the SNP analysis. From such a low level of sequence diversity we conclude that the 3 patients had been infected with the same strain and that Ml6-50 is ancestral to Ml2-10 and Ml6-55. However, the chronology of diagnosis in these patients does not fit with the ancestry of the SNPs, since patient Ml2-10 developed leprosy a few years before the other 2 cases. While it is possible that the 3 patients contracted the disease from different sources, there is a theoretical possibility that Ml2-10 contracted leprosy from either Ml6-50 or, more probably, from his/her sibling Ml6-55 before the onset of symptoms in the transmitter. This is an important notion because the mechanism of leprosy transmission is still obscure, and it is unknown at which stage of the disease a person becomes infectious.

Previous transmission studies of human and animal reservoirs of *M. leprae* showed a low level of sequence diversity between strains from the same geographical area (<3 SNPs) [15]. With the high resolution of whole-genome sequencing applied in this study, we unambiguously confirmed that the same dapsone resistant strain was the causative agent of 3 leprosy cases in the area of Kankan in Guinea-Conakry.

Overall, our findings show, for the first time, cases of primary dapsone and rifampicin resistance in West Africa. Although dapsone resistance is not considered a major threat for leprosy control, such cases are a good epidemiological indicator of active transmission of the disease. In addition, our results highlight the interest for a pilot study on adherence to MDT in endemic countries. National programs frequently use selftreatment by the patient with the assistance of families when supervised treatment is not possible. It is likely in the case of self-treatment that adherence is less effective than when patients are supervised by healthcare workers. Finally, the discovery of rifampicin-resistant strains, like the one reported here, is a bigger concern due to the limited alternatives in leprosy treatment of rifampicin-resistant cases.

Supplementary Data

Supplementary materials are available at http://cid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

Acknowledgements. We acknowledge the patients and staff for their invaluable participation and support.

Financial support. This work was supported by the Fondation Raoul Follereau.

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. WHO Expert Committee on Leprosy, World Health Organization. Eighth report. Geneva, 2010.

2. WHO. Weekly epidemiological record: Global leprosy update, 2014: need for early case detection. 2015: 431–76.

3. Matsuoka M. Global surveillance system to monitor the development of drug resistance in *Mycobacterium leprae*. Res Rep Trop Med 2015; 6:75.

4. Guerrero MI, Colorado CL, Torres JF, León CI. Is drug-resistant *Mycobacterium leprae* a real cause for concern?: First approach to molecular monitoring of multibacillary Colombian patients with and without previous leprosy treatment. Bioméd Rev Inst Nac Salud 2014; 34(suppl 1):137–47.

5. Singh P, Busso P, Paniz-Mondolfi A, et al. Molecular drug susceptibility testing and genotyping of *Mycobacterium leprae* strains from South America. Antimicrob Agents Chemother 2011; 55:2971–3.

6. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30:2114–20.

7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie2. Nat Methods 2012; 9:357–9.

8. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012; 22:568–76.

9. Monot M, Honoré N, Garnier T, et al. Comparative genomic and phylogeographic analysis of Mycobacterium leprae. Nat Genet 2009; 41:1282–9.

10. Zhang L, Budiawan T, Matsuoka M. Diversity of potential short tandem repeats in Mycobacterium leprae and application for molecular typing. J Clin Microbiol 2005; 43:5221–9.

11. Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern Mycobacterium leprae. Science 2013; 341:179–83.

12. Williams DL, Pittman TL, Gillis TP, Matsuoka M, Kashiwabara Y. Simultaneous detection of Mycobacterium leprae and its susceptibility to dapsone using DNA heteroduplex analysis. J Clin Microbiol 2001; 39:2083–8.

13. Matsuoka M, Budiawan T, Aye KS, et al. The frequency of drug resistance mutations in Mycobacterium leprae isolates in untreated and relapsed leprosy patients from Myanmar, Indonesia and the Philippines. Lepr Rev 2007; 78:343–52.

14. Singh P, Benjak A, Schuenemann VJ, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of Mycobacterium lepromatosis. Proc Natl Acad Sci 2015; 112:4459–64.

15. Sharma R, Singh P, Loughry WJ, et al. Zoonotic leprosy in the southeastern United States. Emerg Infect Dis 2015; 21:2127–34.

Table S1: Allelic diversity of microsatellite loci in strains Ml2-10, Ml6-50 and Ml6-55 compared to S13 [1] and the previously described strains Thai-53, Airaku-3 and Zensho-4 [2] – U: undetermined; red shows VNTR associated with SNP subtype 4N; bold font shows VNTR that differ.

	Genotype	$(T)^{8}(A)^{6}$	$(T)^{6}(N)^{7}(T)^{8}$	$(A)^{9}$	$(G)^{9}$	$(C)^{9}$	$(G)^{10a}$	$(G)^{10b}$	$(G)^{11}$	$(G)^{12}$	$(C)^{16}(G)^{8}$	$(C)^{20}$	$(G)^{22}$	$(CG)^{6}$	$(AC)^{8a}$	$(AC)^{8b}$	$(AC)^{9}$
Ml2-10	4N	8 & 6	6 & 7 & 8	8	9	8	9	9	**9**	**10**	U	**12**	U	6	9	7	7
Ml6-50	4N	8 & 6	6 & 7 & 8	8	9	8	9	9	**10**	**10**	U	**11**	U	6	9	7	7
Ml6-55	4N	8 & 6	6 & 7 & 8	8	9	8	9	9	**10**	**9**	U	**11**	U	6	9	7	7
S13	4N	8 & 6	6 & 7 & 8	8	10	8	11	9	**11**	**9**	U	**U**	U	6	9	7	8
Thai-53	1A	8 & 6	6 & 7 & 8	9	9	10	12	11	10	9	14 & 8	9	17	6	11	7	7
Airaku-3	1D	8 & 6	6 & 7 & 8	8	9	10	9	11	11	10	16 & 8	13	18	6	10	6	8
Zensho-4	3K	8 & 6	6 & 7 & 8	8	9	10	11	12	11	9	17 & 7	14	14	6	8	8	7

Table S1: Continued

$(CA)^{6}$	$(TA)^{9}$	$(TA)^{10}$	$(TA)^{13}$	$(AT)^{10}$	$(AT)^{15}$	$(AT)^{17}$	$(TA)^{18}$	$(ACC)^{5}$	$(GGT)^{5}$	$(AGT)^{5a}$	$(ACT)^{5}$	$(GAT)^{9}$	$(AGA)^{20}$	$(CACGG)^{3}$
6	8	8	U	6	U	U	U	5	4	5	U	9	**13**	3
6	8	8	U	6	U	U	U	5	4	5	U	9	**12**	3
6	8	8	U	6	U	U	U	5	4	5	U	9	**12**	3
6	7	8	U	7	U	U	U	5	4	5	U	U	**U**	3
6	9	11	23	8	13	13	21	5	5	5	5	11	14	3
6	10	7	21	7	18	13	17	5	4	5	5	11	13	3
6	8	10	17	9	20	13	15	5	4	5	5	13	10	3

Table S2: Allelic diversity of minisatellite loci in strains Ml2-10, Ml6-50 and Ml6-55 compared to S13 and previously described strains Thai-53, Airaku-3 and Zensho-4 (2) – U: undetermined, red shows VNTR associated with SNP subtype 4N

	6-3a	6-3b	6-7	7-3	10-4	12-5	15-3	18-8	21-3	23-3	27-5
Ml2-10	3	3	6	U	4	U	3	U	2	2	5
Ml6-50	3	3	6	U	4	U	3	U	2	2	5
Ml6-55	3	3	6	U	4	U	3	U	2	2	5
S13	3	3	6	U	4	U	3	U	2	2	5
Thai-53	3	3	6	3	4	4	3	8	3	2	5
Airaku-3	3	3	6	3	4	4	3	8	1	2	5
Zensho-4	4	3	9	3	4	3	3	7	3	2	5

Chapter 4.3 – Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases

Mariane M. A. Stefani[1*#], Charlotte Avanzi[2*], Samira Bü hrer-Sé kula[1,3,4], Andrej Benjak[2], Chloé Loiseau[2,5], Pushpendra Singh[2,6], Maria A. A. Pontes[7], Heitor S. Gonç alves[7], Emerith, M. Hungria[1], Philippe Busso[2], Jé ré mie Piton[2], Maria I. S. Silveira[7], Rossilene Cruz[3,8], Antô nio Schetinni[8], Maurí cio B. Costa[9], Marcos C. L. Virmond[10], Suzana M. Diorio[10], Ida M. F. Dias-Baptista[10], Patricia S. Rosa[10], Masanori Matsuoka[11,12], Maria L. F. Penna[13], Stewart T. Cole[2#], Gerson O. Penna[14,15#]

[1]Tropical Pathology and Public Health Institute, Federal University of Goiás, Goiania, Goiás, Brazil
[2] Global Health Institute, É cole Polytechnique Fé dé rale de Lausanne, Switzerland
[3] University of Amazonas State, Manaus, Amazonas, Brazil
[4] Tropical Medicine Foundation Dr. Heitor Vieira Dourado, Manaus, Amazonas, Brazil
[5] Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Switzerland
[6] Department of Microbiology and Biotechnology Centre, Maharaja Sayajirao University of Baroda, Vadodara, India
[7] Dona Libânia Dermatology Centre, Fortaleza, Ceará, Brazil
[8] Tropical Dermatology and Venerology, Alfredo da Matta Foundation, Manaus, Amazonas, Brazil
[9] Faculty of Medicine, Federal University of Goiás, Goiania, Goiás, Brazil
[10] Lauro Souza Lima Institute, Bauru, São Paulo, Brazil
[11] Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Higashimurayama, Tokyo, Japan
[12] Jyu-kanbo National Museum, Kusatsu, Gunma, Japan
[13] Epidemiology and Biostatistics Department, Universidade Federal Fluminense, Rio de Janeiro, Brazil
[14] Tropical Medicine Centre, University of Bras´ılia, Bras´ılia DF, Brazil
[15] Fiocruz, Brasilia, Brazil

2017. PNTD. 11(6): e0005598

[*] These authors contributed equally to this work
[#]Corresponding author(s)

Contributions: DNA extraction, library preparation, array capture, interpretation of results, manuscript preparation

This is a pre-print, author-modified version of an article accepted for publication in Plos Neglected tropical disease following peer review. The version of record is available online in free open access at: **doi.org/ 10.1371/journal.pntd.0005598**

ABSTRACT

BACKGROUND

Since leprosy is both treated and controlled by multidrug therapy (MDT) it is important to monitor recurrent cases for drug resistance and to distinguish between relapse and reinfection as a means of assessing therapeutic efficacy. All three objectives can be reached with single nucleotide resolution using next generation sequencing and bioinformatics analysis of *Mycobacterium leprae* DNA present in human skin.

METHODOLOGY

DNA was isolated by means of optimized extraction and enrichment methods from samples from three recurrent cases in leprosy patients participating in an open-label, randomized, controlled clinical trial of uniform MDT in Brazil (U-MDT/CT-BR). Genome-wide sequencing of *M. leprae* was performed and the resultant sequence assemblies analyzed *in silico*.

PRINCIPAL FINDINGS

In all three cases, no mutations responsible for resistance to rifampicin, dapsone and ofloxacin were found, thus eliminating drug resistance as a possible cause of disease recurrence. However, sequence differences were detected between the strains from the first and second disease episodes in all three patients. In one case, clear evidence was obtained for reinfection with an unrelated strain whereas in the other two cases, relapse appeared more probable.

CONCLUSIONS/SIGNIFICANCE

This is the first report of using *M. leprae* whole genome sequencing to reveal that treated and cured leprosy patients who remain in endemic areas can be reinfected by another strain. Next generation sequencing can be applied reliably to *M. leprae* DNA extracted from biopsies to discriminate between cases of relapse and reinfection, thereby providing a powerful tool for evaluating different outcomes of therapeutic regimens and for following disease transmission.

AUTHOR SUMMARY

Leprosy, one of the most ancient human infectious diseases, affects skin and nerves and is caused by *Mycobacterium leprae* infection. Despite the effective use of multidrug therapy/MDT since the 80´s, over 200,000 new cases are reported yearly, indicating active transmission, especially in India and Brazil. Although rare, recurrent clinical manifestations after MDT can occur due to leprosy reactions, relapse by drug resistance, insufficient treatment or reinfection. Relapse and reinfection cannot be differentiated clinically and molecular genotyping of a predefined set of loci have limited resolution due to exceptional *M. leprae* genome conservation and low sequence diversity between strains from the same geographical area. This is the first report that has compared whole genome sequences of *M. leprae* strains from original and recurrent leprosy episodes. *M. leprae* genome differences were detected between the strains from the first and second episodes in the three patients. In one patient, there was clear evidence for reinfection with an unrelated strain whereas the other two were considered true relapses due to minor strain differences. No known drug resistance mutations were detected, excluding drug resistance as the recurrence cause. Next generation sequencing of *M. leprae* DNA discriminates relapse from reinfection representing a powerful tool for evaluating different disease outcomes and transmission.

Introduction

Leprosy is a complex dermato-neurologic and systemic disease [1] primarily caused by *Mycobacterium leprae* or to a much lesser extent by *Mycobacterium lepromatosis*.[2] Despite a strong decrease in leprosy prevalence since the systematic implementation of multidrug therapy (MDT) in the 1980's, the incidence of disease, the major indicator of active transmission, remains

high in many countries, especially in India and Brazil, showing that transmission continues unabated.[3] Overall, more than 200,000 new leprosy cases are reported each year worldwide.[3]

The MDT regimen for leprosy consists of different antibiotic combinations that are prescribed based on the number of skin lesions: a six-month regimen of rifampicin and dapsone for paucibacillary (PB) patients (<5 skin lesions) and a twelve month regimen of rifampicin, dapsone and clofazimine for multibacillary (MB) patients (>5 skin lesions).[4] In 2002, WHO proposed that a uniform MDT regimen (U-MDT) should be considered to treat all types of leprosy in order to facilitate leprosy control. In 2007, an open-label randomized and controlled clinical trial (uniform multidrug therapy for leprosy patients in Brazil, U-MDT/CT-BR) was initiated to compare U-MDT with the regular MDT for PB and MB patients.[5, 6] Clinical monitoring is still taking place with special emphasis on disease recurrence and leprosy type 1 and type 2 reactions (T1R/T2R).

An increased relapse rate and the possible emergence of drug resistance are major concerns for the shortened MDT proposal for MB patients. It is therefore important to address this issue by analyzing in depth all recurrent cases from the U-MDT/CT-BR trial. Molecular genotyping techniques, such as typing selected single nucleotide polymorphisms (SNP) or counting variable number tandem repeats (VNTR) have been used to differentiate reinfection from relapse.[7–11] However, the resolution of such techniques is often limited because of the exceptional level of genome conservation in *M. leprae* and the limited sequence diversity between strains from the same geographical area in particular.[12] In contrast, genome-wide approaches provide higher resolution and accuracy compared to genotyping based on a predefined set of loci, but are technically more complex. High throughput sequencing is becoming increasingly efficient and cost-effective with purified DNA but is more challenging with clinical specimens such as DNA extracted directly from skin biopsies, especially from formalin-fixed paraffinembedded (FFPE) samples.

In this study, we investigated three recurrent cases of leprosy from the U-MDT/CT-BR trial to determine whether recurrence was due to drug resistance, bacterial persistence or to reinfection. To achieve this, we compared whole genome sequencing analysis of *M. leprae* collected from skin lesions at the initial diagnosis and during the recurrence of the disease and correlated the sequence data with the clinical, microbiologic and serologic findings.

Methods

Ethics statement

This study was approved by the regional research ethical committees, by the National Committee for Ethics in Research (CONEP, National Health Council/ Ministry of Health, Brazil, protocol # 001/06) and by the human and animal research ethics committee from the Federal University of Goiás (CEMHA/HC/UFG protocol # 166/2011). Written informed consent was obtained from all adult subjects and a parent or guardian of participants under the age of 18 years, provided informed consent on their behalf prior to inclusion in the study (ClinicalTrials.gov identifier: NCT00669643).

Study design

Three recurrent cases of leprosy identified in the U-MDT/CT-BR trial were investigated (**Table 1**). Clinical diagnosis and monitoring were carried out at the National Reference Canter in Ceará state, Northeast Brazil. Leprosy diagnosis was confirmed by bacteriological analysis of slit skin smears and by histopathological examination of biopsies taken from active skin lesions.[6] At the first visit, patients had a complete dermato-neurological examination by a

Table 1. Main demographic and clinical features of recurrent cases from the U-MDT/CT-BR trial.

Patient's identification	1126	3208	2188
Age at diagnosis (years)	32	17	20
Sex	Male	Male	Male
Date of U-MDT (month/year)			
Start	06/2007	10/2007	09/2007
Completion	11/2007	03/2008	02/2008
Relapse date (month/year)	09/2011	04/2015	11/2014
Months between the end of U-MDT and recurrence of symptoms	46	86	79
Ridley-Jopling classification			
First diagnosis	BL	LL	LL
Relapse	LL	LL	LL
Bacillary index			
First diagnosis	4.0	4.75	3.5
Relapse	4.0	4.0	4.2
Lowest BI (months follow-up)			
	1.25	1.0	3.0
	31	43	76
Number of reactions and neuritis during follow up	4	7	3
Clinical evolution			
	Neuritis (n = 1)	Neuritis (n = 1)	
	T2R (n = 2)	T2R (n = 5)	T2R (n = 2)
	T1R (n = 1)	T1R (n = 1)	T1R (n = 1)

CRF: case report form; MDT: multidrug therapy; U-MDT: uniform MDT; T1R: BI: bacilloscopic index; type 1 reaction; T2R: type 2 reaction.

dermatologist with expertise in leprosy diagnosis, when the number and the body distribution of skin lesions and affected nerves were registered. Biopsy of skin lesion, venous blood and skin smear material from six sites for bacilloscopy were collected. During the clinical monitoring, patients attended the established schedule for clinical/laboratory monitoring (monthly appointment during the first year and thereafter, yearly). All patients were advised to return to an urgent appointment at the reference center in case any discomfort or new clinical manifestation appeared. In this study, the following case definitions for leprosy reactions were employed: T1R was defined as an acute clinical manifestation, usually characterized by the exacerbation of pre-existing lesions, or the appearance of new lesions. T2R was characterized by the sudden appearance of tender erythematous skin nodules (erythema nodosum leprosum/ENL) mainly accompanied by fever and other systemic symptoms such as joint pain, bone tenderness, neuritis, edema, malaise, anorexia with or without lymphadenopathy. In the clinical diagnosis of reactions, skin signs were obligatory, nerve and systemic signs were noncompulsory while neuritis, malaise, and fever could be present in both types of reaction. Treatment for leprosy reactions followed the guidelines from the Brazilian Ministry of Health.

Patients with clinical manifestations not fulfilling these previously described criteria were considered suspected cases of relapses and were clinically examined by the assistant dermatologist, by the PI (GOP) and by an expert member of the independent steering committee (Dr. Sinesio Talhari). Additionally, in these patients skin smears and biopsies were collected from new lesions and used to investigate drug susceptibility (inoculation in BALB/c mice, sequencing of the *rpoB*, *folP1*, *gyrA* and *gyrB* genes and whole genome sequencing).

As part of the U-MDT/CT-BR trial, a well-prepared biobank of biopsies from leprosy skin lesions and serum samples, collected at various time points during treatment and monitoring, was assembled and has been properly maintained at recruitment sites and an extra back-up has been kept at the coordination center. For this study, we used skin biopsies from the first episode that were formalin-fixed and paraffin-embedded to allow long-term storage and serum samples collected at diagnosis and at various time-points during

and after treatment (**Table 1 and S1 Table**). Serum IgM antibodies to *M. leprae*-specific PGL-1 antigen (0·01µg/ mL NT-P-BSA) and serum IgG antibodies to the synthetic LID-1 (1µg/mL LID-1) antigen were detected by enzyme-linked immunosorbent assay (ELISA).[13, 14]

Patients showing recurrent symptoms after treatment had biopsies taken from new lesions (**Table 1 and S1 Table**), which were used as the source of *M. leprae* for drug susceptibility testing in BALB/c mice [15] (treated with dapsone, rifampicin or no drug) and for partial [16] and whole genome sequencing.

DNA extraction from tissue

A truXTRAC™ FFPE DNA kit (Covaris) was used following the manufacturer's recommendation with some optimization. Briefly, ten 20µm FFPE tissue sections for each sample were pooled in a screw-cap microTUBE in duplicate or triplicate. Paraffin was removed and the tissue rehydrated with 100µl of tissue SDS buffer using a focused-ultrasonicator series S2 with the following settings: intensity = 5, cycles per burst = 200, time = 300s, temperature = 20˚C. Digestion was done using a 40µl mixture of proteinase K (20 mg/ml) and lysozyme (10 mg/ml) using a focused-ultrasonicator with the same settings as above except for the time set at 10s.

Digestion occurred at 56˚C overnight followed by 1 h at 80˚C to reverse the formaldehyde crosslinks. Finally, DNA was isolated from lysates using the columns of the truXTRAC FFPE DNA kit and eluted in 50µl of Covaris BE buffer. DNA was quantified using a Qubit fluorometer (ThermoFisher). For samples 1126–2011 and 2188–2014, which had been passaged in mice, DNA was extracted from mouse footpad suspensions then sheared to ~600 bp by ultrasonication and purified with AMPure beads, before library preparation.

The quantity of DNA was assessed after each critical step i.e. DNA extraction, library preparation and amplification post-array capture (**S2 Table**). Since the quality of DNA is known to be low after FFPE extraction, we did not fragment the DNA with the Covaris method as it was already fragmented nor did we size select our libraries to avoid losing too much DNA.

Library preparation and sequencing

DNA from each extract was used to prepare Illumina libraries using a Kapa Hyper Prep kit (Kapa Biosystem) as described elsewhere.[17] To remove host DNA from the libraries, we used a custom-synthesized oligonucleotide array (Agilent) spanning the entire *M. leprae* genome.[18] Quality of the captured and re-amplified library was assessed using the Fragment Analyzer system (Advances Analytical technologies, Inc). The size of the captured library was 180bp and the concentration 52ng/µl. Sequencing was performed on an Illumina Hi-Seq 2500 instrument.

Sequence analyses

Raw reads from the same sample were merged and processed as described elsewhere [17] by adapterand quality-trimming and alignment with the *M. leprae* TN reference genome (NCBI a.n. AL450380.1). To avoid false positive SNP calls the following cutoffs were applied: minimum overall coverage of 5 non-duplicated reads, minimum of 3 non-duplicated reads supporting the SNP, mapping quality score greater than 8, base quality score greater than 15 and a SNP frequency above 80%.

SNPs and short insertions and deletions (InDels) were compared between index and second episodes for each recurrent case. Unique sets of SNPs for each genome were established by comparison with the list of SNPs from 20 *M. leprae* genomes published elsewhere (**S3 Table**). [9, 18, 19] All unique and/or discriminatory variants were manually visualized using the IGV browser [20] to

check for possible alignment inconsistencies. We additionally genotyped all samples using the SNP model described in Monot *et al*. and inferred *in silico* the VNTR copy number for 33 out of 44 known VNTR loci (11 loci were too large to be spanned with Illumina reads). [9, 11, 21]

Results

Demographics and diagnosis

The U-MDT/CT-BR study initially enrolled 858 patients of whom 78.4% were classified as MB. During follow-up, four of the treated patients presented with new symptoms between four and eight years after completion of U-MDT and three of these were re-investigated in this study. These participants were three young male leprosy patients (# 1126, 2188 and 3208) from Fortaleza, Ceará, Northeast Brazil, an endemic city for leprosy. The main clinical and laboratory characteristics of these three patients with recurrent signs of leprosy after U-MDT are shown in **Table 1**. In all three cases, leprosy was first diagnosed in 2007 but the patients displayed new clinical signs, which were not associated with leprosy reactions, between 2011 and 2015.

In these three patients, original leprosy skin lesions detected at diagnosis, disappeared after specific treatment and upon suspicion of relapse/reinfection, new skin lesions were observed in previously unaffected body areas. The timelines of clinical events presented by these patients during follow up (**S1 Fig**) highlight their high propensity to develop leprosy reactions, especially T2R, although all of them also developed T1R. These records also demonstrate that leprosy reactions and relapse/reinfection occurred at different time points. The timelines also illustrate the evolution of bacilloscopic index (BI) during follow up. In one case, the BI at the second episode was higher than the BI at the first episode.

In addition, the first diagnosis revealed that the three MB patients showed high IgM and IgG antibody levels to PGL-1 and LID-1 antigens, respectively (**S1 Fig**). Since these biomarkers have been used to monitor the disease state, we measured antibody levels by ELISA before, during and after U-MDT. Overall, the antibody titers gradually declined but remained above the threshold for positivity for at least one of the antigens during the study period except for patient 1126. This patient showed an antibody titer below the threshold just before the recurrence of the disease (39 months after U-MDT) and then both antibody titers increased by the time of recurrent disease. By contrast, despite oscillating levels of PGL-1 antibody for 3208,

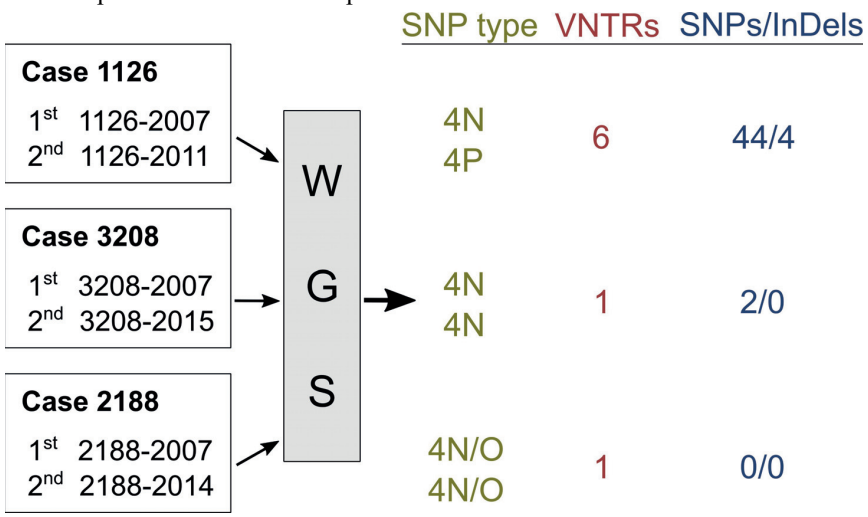


Figure 1: Summary of the whole genome sequencing analysis (WGS) between the *M. leprae* strains investigated at the 1st and 2nd disease occurrences.

antibody titers, especially to LID-1, remained high for 3208 and 2188 during the entire study period.

Drug susceptibility results

M. leprae from the recurrent lesions (1126–2011, 2188–2014) was inoculated into mice and only multiplied in the untreated animals, indicating that the bacilli were viable but susceptible to dapsone and rifampicin. It was not possible to inoculate mice with the sample from 3208– 2015. Analysis of the *rpoB*, *folP1*, *gyrA* and *gyrB* genes revealed a wild-type sequence in all six strains, confirming susceptibility to rifampicin, dapsone, and fluoroquinolones, respectively, in all cases.

Whole-genome analysis

Sufficient whole genome read coverage was obtained from the six *M. leprae* samples for genotyping and comparative genomic analyses (**S4 Table**).

The recurrent strain 1126–2011 was clearly distinct from the primary strain 1126–2007, and differed in 44 SNPs, 4 InDels and 6 VNTR loci (**Fig 1 and S5 Table**). Furthermore, 1126– 2007 and 1126–2011 share no SNPs that might indicate close relatedness or direct ancestry.

Strains 3208–2007 and 3208–2015 differed in only two SNPs and one VNTR locus (**Fig 1 and S5 Table**). Both SNPs (T1740863C in an intergenic region and C1803024T in a pseudogene) were present in 3208–2015, indicating that 3208–2015 was certainly the direct progeny of 3208–2007. In addition, eight unique variant nucleotides were restricted to these two samples (compared to the SNPs from 20 previously published *M. leprae* genomes [9, 18, 19] and those from this study), confirming the identity of the strains (**S6 Table**). Interestingly, a cluster of three SNPs leads to missense mutations, in codons 495 and 496 of *asn1*, encoding an Lasparagine permease, which contributes to virulence in *Mycobacterium tuberculosis* [22]. Analysis of 2188–2007 and 2188–2014 revealed identical genome sequences (**Fig 1**). Curiously, both strains belong to a new SNP subtype intermediate between subtypes 4N and 4O. The only difference between the two genomes was found in the (GTA)9 VNTR locus (**S5 Table**), which harbored 11 repeats in 2188–2007 and 12 repeats in 2188–2014. Genome comparisons revealed that both strains share 28 unique variant nucleotides (**S7 Table**). Among them are two missense mutations in *ML0411*, encoding a PPE protein and in *ribD* (*ML1340*), the riboflavin biosynthesis protein. An insertion of 9 nucleotides (GGACATCTA at position 1,219,061) was found in *ML1052*, a putative PucR-like transcriptional regulator, which leads to a modification of the protein. Interestingly this mutation was present at only 30% frequency in 2188–2007, while it was fixed in 2188–2014.

Another frame-shift arising from a dinucleotide insertion was found in *ML0825c*, the ortholog of *rv2358* in *M. tuberculosis* that codes for the protein SmtB, a zinc-sensing transcriptional regulator and member of the AsrR/SmtB family.[23, 24]. The C-terminal part of SmtB is essential for the protein dimerization, zinc binding and DNA recognition. Furthermore, a specific histidine residue (H138 in ML0825c and H117 in *Synechococcus* StmB) is important for the allosteric coupling of the zinc and DNA binding sites in the protein.[25] Modeling of *M. leprae* StmB *in silico* (**S2 Fig**) showed that the frame-shift leads to loss of H138 and should thus impair protein function.

Discussion

The relapse rate is considered to be the most important indicator of the efficacy of a given MDT. On the other hand, reinfection is an indicator of active transmission and the susceptibility of leprosy convalescents to new infections.

This investigation provided a unique opportunity to apply high-resolution whole-genome tools to differentiate relapse from reinfection and to evaluate the impact of U-MDT on antibody levels to two *M. leprae* antigens. MDT affects both cellular and humoral *M. leprae* specific immunity. In MB patients, there is a decline in antibody levels during MDT and patients remain unable to mount a protective Th1 type immunity to *M. leprae* after treatment.[26] Levels of antibodies to PGL-1 and LID-1 were high in all three MB cases at diagnosis, then declined during and after treatment but nonetheless remained above the cut-off point for positivity, especially antibodies to LID-1. Our data is in accordance with previous studies showing decay in antibody titers while sero-reversion is rare in leprosy patients after regular MDT [26, 27]. By the time of recurrent disease, the antibody titers to at least one of the antigens had risen. In our study, patients were carefully monitored for treatment compliance and all completed the U-MDT treatment.

In our study the three MB patients had several episodes of leprosy reactions during follow up including T1R and mainly T2R, in accordance with the reports showing increased propensity of MB patients to develop reactions. [28, 29] In fact, several studies have shown that in some endemic areas the occurrence of T1R in BL/LL patients is higher than T2R. A study about risk factors for leprosy reactions in patients from three endemic countries (Philippines, Nepal, Brazil) showed that among all LL and BL patients, T1R was more frequent than T2R. Another study from Thailand showed that T2R was slightly more prevalent than T1R in lepromatous patients. [31] T1R primarily affects immunologically unstable borderline patients (BL, BT, BB), while although sporadic, it also occurs in LL patients. T1R is characterized by an increased inflammatory Th1-type cell-mediated immunity in pre-existing skin lesions and systemically, in serum and in circulating leukocytes. The capacity of BL/LL patients, who have a predominant Th2 response, to develop T1R was elucidated by studies showing leukocytes with a Th0 profile that produce IFNγ, IL2 and IL4 or a polarized shift to Th1 type response with IFNγ and IL-12p40 mRNA in lesional skin and in leukocytes. [32, 33]

In both leprosy and tuberculosis, host genetic factors and immunological mechanisms determine the outcome of infection so that susceptibility varies among individuals. Case 1126 was unambiguously identified as reinfection because of the extensive polymorphisms between the two strains. Reinfection has long been suspected as a cause of new leprosy episodes and it has been suggested that individuals who have already had leprosy are more likely to be reinfected after treatment due to their inherent immunogenetic susceptibility.[34–36] Around 30% of relapse cases in Recife, northeast Brazil, were reported to be in contact with other leprosy patients and more often from the same family or household.[8] Leprosy case 1126 is an example of "family disease", because both of the patient's parents had leprosy around five years before his diagnosis, his daughter and partner had PB leprosy and the partner's cousin, who lives in the same household, was diagnosed with MB leprosy but failed to complete MDT due to alcohol addiction.

The extremely limited genomic variability detected between strains from the same geographical origin poses a challenge in distinguishing between relapse or reinfection with a closely related strain. In a recent paper, Avanzi *et al*. showed that a strain infecting three patients in the same region of Guinea Conakry differed in only two SNPs [17] and four VNTRs. In our study, two SNPs and one polymorphic VNTR were found. While individual VNTRs carry virtually no ancestral information due to the risk of homoplasy and mutation reversion, the fact that only two SNPs were found

strongly indicates that the recurrent strain was directly derived from the original infection. It should be recalled that in our study skin biopsies were taken from two different lesions in different body areas. Likewise, in the case of 2188 only one polymorphic VNTR locus distinguished between the first and the recurrent infection, and the absence of SNPs confirms the strain's identity. Furthermore, there was no history of leprosy in either patient 3208's or 2188's households, and both patients had high antibody titers during the study period suggesting continued immunological stimulation by bacterial antigens after treatment. Therefore, based on the genomic analysis, the patients' epidemiologic history and serological data, we consider that the recurrence of leprosy in both patients 3208 and 2188 was due to relapse.

Leprosy presents a variable incubation period which can range from 2–15 years. Although more prevalent in adults, leprosy also occurs in children <15 years, with reports of cases in patients younger than 1 year of age [37] indicating at least in children, short incubation period of the disease. However, nothing is known about the incubation period of reinfection, especially in genetically susceptible individuals who remain exposed to the bacilli in endemic areas. In this study, the reinfection case was observed 4 years after the conclusion of treatment, indicating a relatively short incubation period but which is in accordance with the reported range of the incubation period of the disease. The availability and larger use of whole genome sequencing studies of *M. leprae* in recurrent leprosy and leprosy reinfection can clarify the duration of incubation period in such cases. Further investigations of other such cases will give us a more definitive picture of characteristics of reinfection.

It is theoretically possible that the original infection in leprosy could involve more than one strain of *M. leprae* and, that the recurrence could be a relapse due the regrowth of one of the sub-populations of *M. leprae*, that had been under-treated by the first course of MDT. However, although possible, in our study this probability was implausible, since in all three patients investigated, including the reinfection case, genomic sequences of the *M. leprae* strains responsible for the original infections showed no mutation associated with drug resistance. Therefore, even if the original infection had involved more than one strain of *M. leprae*, these strains were MDT susceptible.

To conclude, this study is the first to demonstrate that it is possible to differentiate reinfection from relapse in leprosy in a field setting with a follow up period extended to eight years. This provides a proof-of-concept and emphasizes the value of whole genome sequencing in clinical follow up of leprosy. Importantly, the extended observation period allowed identification of relapses/reinfection. *M. leprae* grows very slowly and has a relatively long incubation time, so shorter periods of monitoring would be unlikely to provide sufficient clinical evidence to suspect relapse or reinfection. Also the two relapse cases in this study exemplify the superiority of whole-genome sequencing over genotyping a limited subset of loci or VNTR typing. For instance, the current SNP genotyping scheme can only detect distinct *M. leprae* lineages [9], which is not useful for analyzing closely related strains. VNTRs can distinguish such strains but do not reflect the overall genetic distance (**Fig 1**) nor convey information about strain ancestry. Improvements in sample preparation have made whole-genome sequencing more applicable routinely and we expect that recent technological advances will culminate in sequencing platforms that can be used to deliver whole genome coverage at the point of diagnosis within days of seeing the patient.[38]

All raw sequence read files have been deposited in the trace archive of the National Center for Biotechnology Information Sequence Read Archive under accession no.

SRP078228.

Acknowledgments

We are thankful to the U-MDT/CT-Br staff and patients for their cooperation. We also thank the Genomic Technologies Facility at the University of Lausanne for Illumina sequencing and technical support.

Author Contributions

Conceptualization: MMAS SBS GOP STC.

Data curation: MMAS CA SBS AB CL PS STC MLP GOP.

Formal analysis: MMAS CA SBS AB CL PS STC MLP GOP.

Funding acquisition: GOP STC.

Investigation: MMAS CA SBS AB CL PS EMH PB JP AS MBC MCLV SMD.

Project administration: GOP STC.

Resources: GOP MMAS STC MCLV.

Supervision: GOP MMAS STC.

Validation: GOP MMAS STC MCLV.

Visualization: STC MMAS GOP SBS.

Writing – original draft: MMAS CA SBS AB STC GOP.

Writing – review & editing: MMAS CA SBS AB CL PS MAAP HSG EMH PB JP MISS RC AS MBC MCLV SMD IMFDB PSR MM MLFP STC GOP.

References

1. Britton WJ, Lockwood DN. Leprosy. Lancet. 2004; 363(9416):1209–19. PMID: 15081655. https://doi. org/10.1016/S0140-6736(04)15952-7

2. Han XY, Seo YH, Sizer KC, Schoberle T, May GS, Spencer JS, et al. A new Mycobacterium species causing diffuse lepromatous leprosy. Am J Clin Pathol. 2008; 130(6):856–64. PMID: 19019760. https:// doi.org/10.1309/AJCPP72FJZZRRVMM

3. WHO. Global leprosy update, 2014: need for early case detection. 2015.

4. WHO. Model Prescribing Information: Drugs Used in Leprosy. Geneva1998.

5. Penna ML, Buhrer-Sekula S, Pontes MA, Cruz R, Goncalves Hde S, Penna GO. Primary results of clinical trial for uniform multidrug therapy for leprosy patients in Brazil (U-MDT/CT-BR): reactions frequency in multibacillary patients. Lepr Rev. 2012; 83(3):308–19. PMID: 23356032.

6. Penna ML, Buhrer-Sekula S, Pontes MA, Cruz R, Goncalves Hde S, Penna GO. Results from the clinical trial of uniform multidrug therapy for leprosy patients in Brazil (U-MDT/CT-BR): decrease in bacteriological index. Lepr Rev. 2014; 85(4):262–6. Epub 2015/02/14. PMID: 25675650.

7. Oskam L, Dockrell HM, Brennan PJ, Gillis T, Vissa V, Richardus JH. Molecular methods for distinguishing between relapse and reinfection in leprosy. Trop Med Int Health. 2008; 13(10):1325–6. Epub 2008/ 10/22. https://doi.org/10.1111/j.1365-3156.2008.02134_1.x TMI2134_1 [pii]. PMID: 18937747.

8. da Silva Rocha A, Cunha Dos Santos AA, Pignataro P, Nery JA, de Miranda AB, Soares DF, et al. Genotyping of *Mycobacterium leprae* from Brazilian leprosy patients suggests the occurrence of reinfection or of bacterial population shift during disease relapse. J Med Microbiol. 2011; 60(Pt 10):1441–6. Epub 2011/05/21. https://doi.org/10.1099/jmm.0.029389-0

jmm.0.029389–0 [pii]. PMID: 21596907; PubMed Central PMCID: PMC3347867.

9. Monot M, Honore N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009; 41(12):1282–9. PMID: 19881526. https://doi.org/10.1038/ng.477

10. Zhang L, Budiawan T, Matsuoka M. Diversity of potential short tandem repeats in *Mycobacterium leprae* and application for molecular typing. J Clin Microbiol. 2005; 43(10):5221–9. Epub 2005/10/07. 43/10/ 5221 [pii] https://doi.org/10.1128/JCM.43.10.5221-5229.2005 PMID: 16207987; PubMed Central PMCID: PMC1248435.

11. Kimura M, Sakamuri RM, Groathouse NA, Rivoire BL, Gingrich D, Krueger-Koplin S, et al. Rapid variable-number tandem-repeat genotyping for *Mycobacterium leprae* clinical specimens. J Clin Microbiol. 2009; 47(6):1757–66. Epub 2009/04/24. https://doi.org/10.1128/JCM.02019-08 JCM.02019-08 [pii]. PMID: 19386839; PubMed Central PMCID: PMC2691099.

12. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the southern United States. N Engl J Med. 2011; 364(17):1626–33. PMID: 21524213. https:// doi.org/10.1056/NEJMoa1010536

13. Brett SJ, Payne SN, Gigg J, Burgess P, Gigg R. Use of synthetic glycoconjugates containing the Mycobacterium leprae specific and immunodominant epitope of phenolic glycolipid I in the serology of leprosy. Clin Exp Immunol. 1986; 64(3):476–83. PMID: 2431812.

14. Duthie MS, Goto W, Ireton GC, Reece ST, Cardoso LP, Martelli CM, et al. Use of protein antigens for early serological diagnosis of leprosy. Clin Vaccine Immunol. 2007; 14(11):1400–8. PMID: 17898185. https://doi.org/10.1128/CVI.00299-07

15. Shetty VP, Wakade AV, Ghate S, Pai VV, Ganapati R, Antia NH. Viability and drug susceptibility testing of *M. leprae* using mouse footpad in 37 relapse cases of leprosy. Int J Lepr Other Mycobact Dis. 2003; 71(3):210–7. Epub 2003/11/12. https://doi.org/10.1489/1544-581X(2003)71<210:VADSTO>2.0.CO;2 PMID: 14608816.

16. Kai M, Nguyen Phuc NH, Nguyen HA, Pham TH, Nguyen KH, Miyamoto Y, et al. Analysis of drug-resistant strains of *Mycobacterium leprae* in an endemic area of Vietnam. Clin Infect Dis. 2011; 52(5):e127– 32. Epub 2011/02/05. https://doi.org/10.1093/cid/ciq217 ciq217 [pii]. PMID: 21292655.

17. Avanzi C, Busso P, Benjak A, Loiseau C, Fomba A, Doumbia G, et al. Transmission of Drug-Resistant Leprosy in Guinea-Conakry Detected Using Molecular Epidemiological Approaches. Clin Infect Dis. 2016; 63(11):1482–4. Epub 2016/08/26. ciw572 [pii] https://doi.org/10.1093/cid/ciw572 PMID: 27558568.

18. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jager G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013; 341(6142):179–83. https://doi. org/10.1126/science.1238286 PMID: 23765279.

19. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proc Natl Acad Sci U S A. 2015; 112(14):4459–64. Epub 2015/04/02. https://doi.org/10.1073/pnas.1421504112 1421504112 [pii]. PMID: 25831531; PubMed Central PMCID: PMC4394283.

20. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011; 29(1):24–6. Epub 2011/01/12. https://doi.org/10.1038/nbt.1754 nbt.1754 [pii]. PMID: 21221095; PubMed Central PMCID: PMC3346182.

21. Fontes AN, Sakamuri RM, Baptista IM, Ura S, Moraes MO, Martinez AN, et al. Genetic diversity of *Mycobacterium leprae* isolates from Brazilian leprosy patients. Lepr Rev. 2009; 80(3):302–15. Epub 2009/12/08. PMID: 19961103.

22. Gouzy A, Larrouy-Maumus G, Wu TD,

Peixoto A, Levillain F, Lugo-Villarino G, et al. *Mycobacterium tuberculosis* nitrogen assimilation and host colonization require aspartate. Nat Chem Biol. 2013; 9 (11):674–6. Epub 2013/10/01. https://doi.org/10.1038/nchembio.1355 nchembio.1355 [pii]. PMID: 24077180; PubMed Central PMCID: PMC3856356.

23. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001; 409(6823):1007–11. PMID: 11234002. https://doi.org/10.1038/ 35059006

24. Canneva F, Branzoni M, Riccardi G, Provvedi R, Milano A. Rv2358 and FurB: two transcriptional regulators from *Mycobacterium tuberculosis* which respond to zinc. J Bacteriol. 2005; 187(16):5837–40. Epub 2005/08/04. 187/16/5837 [pii] https://doi.org/10.1128/JB.187.16.5837-5840.2005 PMID: 16077132; PubMed Central PMCID: PMC1196093.

25. Eicken C, Pennella MA, Chen X, Koshlap KM, VanZile ML, Sacchettini JC, et al. A metal-ligand-mediated intersubunit allosteric switch in related SmtB/ArsR zinc sensor proteins. J Mol Biol. 2003; 333 (4):683–95. Epub 2003/10/22. S0022283603011306 [pii]. PMID: 14568530.

26. Freitas AA, Oliveira RM, Hungria EM, Cardoso LP, Sousa AL, Costa MB, et al. Alterations to antigenspecific immune responses before and after multidrug therapy of leprosy. Diagn Microbiol Infect Dis. 2015; 83(2):154–61. Epub 2015/08/04. https://doi.org/10.1016/j.diagmicrobio.2015.06.021 S07328893(15)00225-4 [pii]. PMID: 26233487.

27. Duthie MS, Hay MN, Rada EM, Convit J, Ito L, Oyafuso LK, et al. Specific IgG antibody responses may be used to monitor leprosy treatment efficacy and as recurrence prognostic markers. Eur J Clin Microbiol Infect Dis. 2011; 30(10):1257–65. PMID: 21544695. https://doi.org/10.1007/s10096-011-1221-2

28. Antunes DE, Araujo S, Ferreira GP, Cunha AC, Costa AV, Goncalves MA, et al. Identification of clinical, epidemiological and laboratory risk factors for leprosy reactions during and after multidrug therapy. Mem Inst Oswaldo Cruz. 2013; 108(7):901–8. https://doi.org/10.1590/0074-0276130222 PMID: 24271045; PubMed Central PMCID: PMC3970646.

29. Kumar B, Dogra S, Kaur I. Epidemiological characteristics of leprosy reactions: 15 years experience from north India. Int J Lepr Other Mycobact Dis. 2004; 72(2):125–33. PMID: 15301592. https://doi.org/10.1489/1544-581X(2004)072<0125:ECOLRY>2.0.CO;2

30. Scollard DM, Martelli CM, Stefani MM, Maroja Mde F, Villahermosa L, Pardillo F, et al. Risk factors for leprosy reactions in three endemic countries. Am J Trop Med Hyg. 2015; 92(1):108–14. Epub 2014/12/

31. 03. https://doi.org/10.4269/ajtmh.13-0221 ajtmh.13-0221 [pii]. PMID: 25448239; PubMed Central PMCID: PMC4347363.

32. Scollard DM, Smith T, Bhoopat L, Theetranont C, Rangdaeng S, Morens DM. Epidemiologic characteristics of leprosy reactions. Int J Lepr Other Mycobact Dis. 1994; 62(4):559–67. Epub 1994/12/01. PMID: 7868954.

33. Sreenivasan P, Misra RS, Wilfred D, Nath I. Lepromatous leprosy patients show T helper 1-like cytokine profile with differential expression of interleukin-10 during type 1 and 2 reactions. Immunology. 1998; 95 (4):529–36. PMID: 9893041.

34. Verhagen CE, Wierenga EA, Buffing AA, Chand MA, Faber WR, Das PK. Reversal reaction in borderline leprosy is associated with a polarized shift to type 1-like *Mycobacterium leprae* T cell reactivity in lesional skin: a follow-up study. J Immunol. 1997; 159(9):4474–83. Epub 1997/10/31. PMID: 9379047.

35. Pannikar V, Jesudasan K, Vijayakumaran P, Christian M. Relapse or late reversal reaction? Int J Lepr Other Mycobact Dis. 1989; 57(2):526–8. Epub 1989/06/01. PMID: 2664046.

36. Rafi A, Donoghue HD, Stanford JL. Application of polymerase chain reaction for the detection of Mycobacterium leprae DNA

in specimens from treated leprosy patients. Int J Lepr Other Mycobact Dis. 1995; 63(1):42–7. Epub 1995/03/01. PMID: 7730718.

37. Brito M, Ximenes R, Gallo M. Retreatment of leprosy relapse. An Bras Dermatol. 2005; 80(3):255–60.

38. Oliveira MB, Diniz LM. Leprosy among children under 15 years of age: literature review. An Bras Dermatol. 2016; 91(2):196–203. Epub 2016/05/19. https://doi.org/10.1590/abd1806-4841.20163661 S0365-05962016000200196 [pii]. PMID: 27192519; PubMed Central PMCID: PMC4861567.

39. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. Genome Med. 2015; 7:99. Epub 2015/09/30. https://doi.org/10.1186/s13073-015-0220-9 [pii]. PMID: 26416663; PubMed Central PMCID: PMC4587849.

Supporting information

S1 Table. Patient samples used for sequencing whole genome of *M. leprae* strains. FFPE: formalin fixed paraffin embedded skin biopsy—MFP: mouse footpad bacilli suspension.

S2 Table. DNA quantification after DNA extraction from the FFPE biopsy samples and MFP samples and after library preparation. FFPE: formalin-fixed paraffin-embedded; MFP:mouse foot pad; LOD: limit of detection.

S3 Table. List of 20 *M. leprae* genomes used to infer unique SNPs in recurrent cases. SNP: single nucleotide polymorphism.

S4 Table. Statistics of *M. leprae* whole genome sequences. [1] Fraction of total reads that aligned to the reference genome TN.

S5 Table. Allelic diversity of VNTR loci in the recurrent leprosy cases. VNTR: variable number tandem repeats; NA: not available because of low coverage at that locus; mixture of multiple alleles; loci where VNTR number varied between the first and second strain are highlighted.

S6 Table. Specific SNPs restricted to 3208–2007 and 3208–2015 strains compared to 24 other *M. leprae* genomes. SNP: single nucleotide polymorphism.

S7 Table. List of 28 unique SNPs in strains 2188–2007 and 2188–2014. SNP: single nucleotide polymorphism. [1]At 30% frequency in 2188–2007.

S1 Fig. *M. leprae* specific anti-phenolic glycolipid 1 (PGL-1) immunoglobulin M (IgM) and anti-LID-1 IgG serology and timelines of clinical events and BI evolution of multibacillary (MB) leprosy patients that had recurrent disease after treatment. Serological data and timelines of clinical events and bacilloscopic index (BI) evolution are depicted in panels A-F: patient # 1126 (A, B); patient # 3208 (C, D); patient # 2188 (E, F). The first serum sample was collected at diagnosis before treatment and sequential samples were collected monthly during treatment and at different times during the follow-up as indicated.

S2 Fig. Structure and polymorphisms in SmtB. Panel A shows the sequence alignment of SmtB homologs. The locations of the α5 metal-binding sites are highlighted in blue and pink. In red is the mutated sequence of SmtB. Panel B shows the structure of the CzrA dimer from *Sthaphylococcus aureus*. Zn, in orange, binds at the interface between the two monomers. Panel C shows a model of the effect of the mutation in ML0825 on the dimer, which compromises the binding of Zn ions. The mutated part is represented in red lines. The protein was modeled using the homology modeling webserver SWISS-MODEL and the structure of the transcriptional repressor CzrA from *Sthaphylococcus aureus* (PDB code 1R1V) as template.

S1 Reference list.

S1 Table: Patient samples used for sequencing whole genome of *M. leprae* strains

Patient #	Type of sample for whole genome sequencing	Year of collection	Sample codes
1126	FFPE	2007	1126-2007
	MFP	2011	1126-2011
3208	FFPE	2007	3208-2007
	FFPE	2015	3208-2015
2188	FFPE	2007	2188-2007
	MFP	2014	2188-2014

FFPE: formalin fixed paraffin embedded skin biopsy MFP: mouse footpad bacilli suspension

S2 Table: DNA Quantification after DNA extraction from the FFPE biopsy samples and MFP samples and after library preparation.

	DNA concentration (ng/µl)	Volume used for library preparation (µl)	Library concentration (ng/µl)
1126-2007_1	<LOD	50	81.7
1126-2007_2	0.058	50	83.4
1126-2007_3	<LOD	50	25.3
1126-2011	<LOD	50	40.9
2188-2007_1	0.411	50	92.9
2188-2007_2	0.318	50	65.1
2188-2007_3	8.64	50	78
2188-2014	0.216	50	54.4
3208-2007_1	0.842	50	129
3208-2007_2	0.136	50	145
3208-2015_1	20.8	50	59.2
3208-2015_2	25.4	42	82.9

FFPE: formalin-fixed paraffin-embedded; MFP: mouse foot pad; LOD: limit of detection.

S3 Table: List of 20 *M. leprae* genomes used to infer unique SNPs in recurrent cases SNP: single nucleotide polymorphism.

Name	Genotype	Described in
TN	1A	[1]
Thai53	1A	[2]
S2	1B	[3]
S11	1D	[3]
3077	2F	[3]
Refshale_16	2F	[3]
SK8	2F	[3]
Jorgen_625	3I	[3]
SK2	3I	[3]
NHDP98	3I	[4]
NHDP55	3I	[4]
I30	3I	[4]
NHDP63	3I	[2]
S9	3K	[3]
Kyoto-2	3K	[5]
S10	3K	[3]
S15	3L	[3]
S13	4N	[3]
S14	4O	[3]
Br4923	4P	[2]
Mx1-22 (*M. lepromatosis*)	Outgroup	[6]

S4 Table: Statistics of *M. leprae* whole genome sequences

	Alignment rate (%)[1]	Average read depth coverage[1]
1126-2007	38.98	6.27
1126-2011	23.47	11.21
3208-2007	42.33	37.19
3208-2015	66.41	31.38
2188-2007	72.36	123.86
2188-2014	16.74	58.25

[1] Fraction of total reads that aligned to the reference genome TN.

S5 Table: Allelic diversity of VNTR loci in the recurrent leprosy cases.

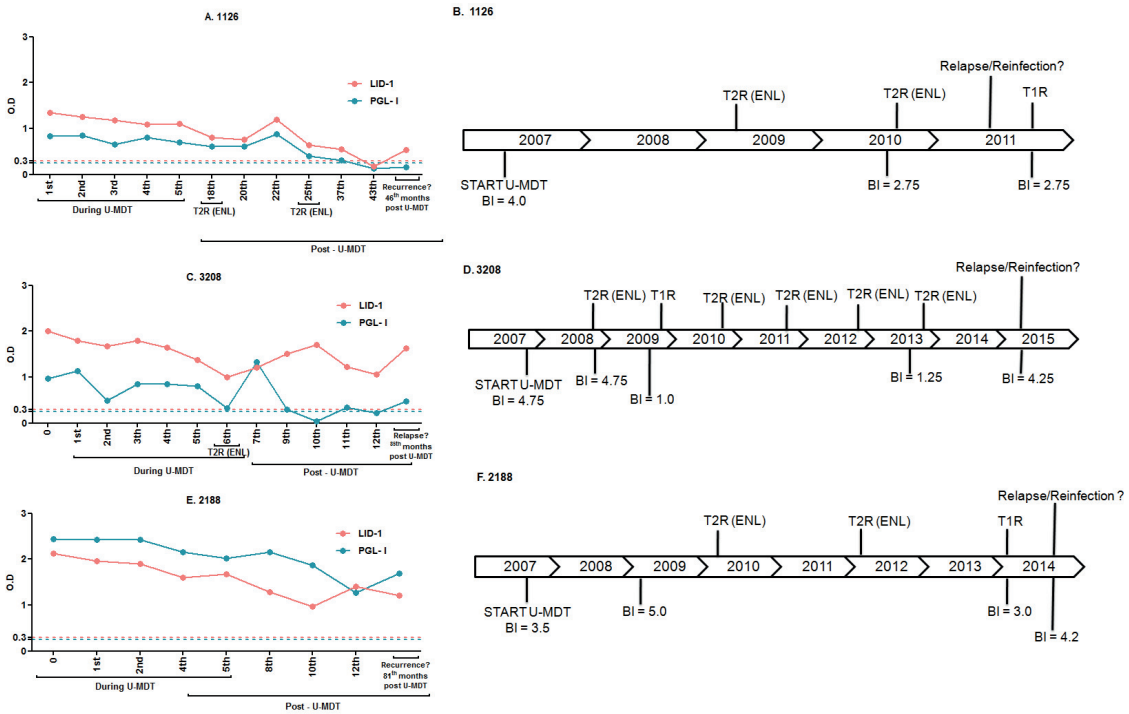
Locus ID	Coordinates	1126		2188		3208	
		1126-2007	1126-2011	2188-2007	2188-2014	3208-2007	3208-2015
(T)8, & (A)6	337466-337473, 337474-337479	8 & 6	8 & 6	8 & 6	8 & 6	8 & 6	8 & 6
(T)6 & (N)7 & (T)8	514181-514186, 514187-514193, 514194-514200	6 & 7 & 8	6 & 7 & 8	6 & 7 & 8	6 & 7 & 8	6 & 7 & 8	6 & 7 & 8
(A)9	1414666-1414674	8	8	8	8	8	8
(G)9	976857-976865	9	10	9	9	9	9
(C)9	2658192-2658200	9	11	9/10	9/10	8/9	8/9
(G)10a	347280-347289	10	9	9	9	13	13
(G)10b	442993-443002	NA	?	9	9	?	9/10
(G)11	1309544-1309554	NA	11	9	9	10/11	12
(G)12	1116443-1116454	9/10	10	9/10	9	11	11
(CG)6	2947291-2947302	6	6	6	6	6	6
(AC)8a	1531185-1531200	9	8	8	8	9	9
(AC)8b	2211035-2211050	7	7/8	7/8	7/8	7	7
(AC)9	1452573-1452590	8	NA	7/8	7/9	8	8
(CA)6	2507097-2507108	6	6	6	6	6	6
(TA)8	3221617-3221632	9	NA	10	10	?	8/9
(TA)9	2844971-2844988	7	6	?	?	7	7
(TA)10	1744091-1744110	8/10	8/9	10/11	10/11	10	10
(AT)10	2951821-2951840	NA	NA	5	5	6	6
(AT)15	948935-948964	NA	NA	?	?	?	?
(AT)17	2597735-2597768	NA	NA	?	13	13	13
(ACC)5	1980049-1980063	5	5	5	5	5	5
(GGT)5	2567251-2567265	4	4	4	4	4	4
(AGT)5a	1237528-1237542	5	5	5	5	5	5
(GTA)9	2583814-2583840	9	10	9/10	11	9	9
(CACCG)3	2562391-2562405	3	3	3	3	3	3
6-3a	1190341-1190358	3	3	3	3	3	3
6-3b	2302531-2302548	3	3	3	3	3	3
6-7	1816857-1816892	6	6	?	8	7	7
7-3	285076-285096	3	3	3	3	3	3
10-4	1139035-1139074	4	NA	4	4	4	4
12-5	1381661-1381723	5	5	5	5	5	5
15-3	2928131-2928175	3	3	3	3	3	3
27-5	687026-687160	5	5	5	5	5	5

VNTR: variable number tandem repeat; NA: not available because of low coverage at that locus; mixture of multiple alleles; loci where VNTR number varied between the first and second strain are highlighted.

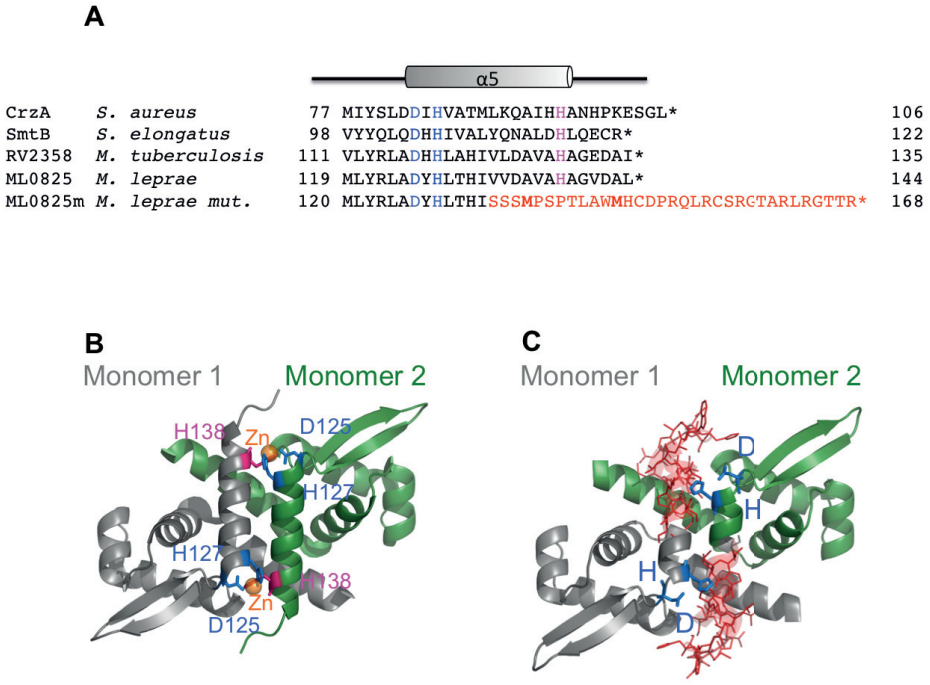
S6 Table: List of 28 unique SNPs in strains 2188-2007 and 2188-2014.

Position	Name	Type	TN base (ref)	Alternate allele	SNP effect	Amino-acid change
509409	*ML0411*	protein_coding	G	A	missense	Asp219Asn
978294	*ML0825*	protein_coding	G	GAT	frameshift	Val128fs
1219061	*ML1052*	protein_coding	G	GGGACATCTA[1]	codon insertion	Leu253-Val255 dup
1599004	*ML1340 / ribD*	protein_coding	C	T	missense	Gly62Asp
14226	*ML0302*	Intergenic	C	CAT	-	
46129	*ML0038*	Pseudogene	C	T	-	
47405	*ML0039*	Pseudogene	C	T	-	
223657	*ML0159*	protein_coding	C	T	synonymous	Ser336Ser
417955	*PPE*	Pseudogene	C	T	-	
428724	-	Intergenic	C	T	-	
534530	*ScoA*	Pseudogene	C	T	-	
657961	-	Intergenic	C	T	-	
676219	-	Intergenic	C	G	-	
809458	-	Intergenic	G	A	-	
1091716	*Ag84*	protein_coding	C	T	synonymous	Leu123Leu
1232532	*ML1068*	Pseudogene	C	T	-	
1320784	*hom*	protein_coding	C	T	synonymous	Ser381Ser
1437154	*ML1214*	protein_coding	G	A	synonymous	Thr165Thr
1450750	-	Intergenic	G	A	-	
1599380	*ML1341*	Pseudogene	C	T	-	
1827911	-	Intergenic	C	CG	-	
2536347	*fprB*	protein_coding	G	A	synonymous	Arg378Arg
2622952	*mce1*	Pseudogene	AC	A	-	
2681117	*ML2256*	Pseudogene	T	TC	-	
2878485	*ML2407*	protein_coding	G	A	synonymous	Pro67Pro
2951819	-	Intergenic	GTATATATATA	G	-	
2972847	*dnaK*	protein_coding	G	A	synonymous	Ser194Ser
3187492	*ML2661*	protein_coding	C	T	synonymous	Ser5Ser

SNP: single nucleotide polymorphism. [1]At 30% frequency in 2188-2007.



S1 Fig: *M. leprae* specific anti-phenolic glycolipid 1 (PGL-1) immunoglobulin M (IgM) and anti-LID-1 IgG serology and timelines of clinical events and BI evolution of multibacillary (MB) leprosy patients that had recurrent disease after treatment. Serological data and timelines of clinical events and bacilloscopic index (BI) evolution are depicted in panels A-F: patient # 1126 (A, B); patient # 3208 (C, D); patient # 2188 (E, F). The first serum sample was collected at diagnosis before treatment and sequential samples were collected monthly during treatment and at different times during the follow-up as indicated.



S2 Fig. Structure and polymorphisms in SmtB. Panel A shows the sequence alignment of SmtBhomologs. The locations of the α5 metal-binding sites are highlighted in blue and pink. In red is the mutated sequence of SmtB. Panel B shows the structure of the CzrA dimer from *Sthaphylococcus aureus*. Zn, in orange, binds at the interface between the two monomers. Panel C shows a model of the effect of the mutation in ML0825 on the dimer, which compromises the binding of Zn ions. The mutated part is represented in red lines. The protein was modeled using the homology modeling webserver SWISS-MODEL[7] and the structure of the transcriptional repressor CzrA from *Staphyloccocus aureus* (PDB code 1R1V) as template.

SI References

1. Cole ST, Eiglmeier K, Parkhill J, et al. Massive gene decay in the leprosy bacillus. Nature 2001;409(6823):1007–11.
2. Monot M, Honoré N, Garnier T, et al. On the origin of leprosy. Science 2005;308(5724):1040–2.
3. Schuenemann VJ, Singh P, Mendum TA, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science 2013;341(6142):179–83.
4. Truman RW, Singh P, Sharma R, et al. Probable Zoonotic Leprosy in the Southern United States. N Engl J Med 2011;364(17):1626–33.
5. Kai M, Nakata N, Matsuoka M, Sekizuka T, Kuroda M, Makino M. Characteristic mutations found in the ML0411 gene of *Mycobacterium leprae* isolated in Northeast Asian countries. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis 2013;19:200–4.
6. Singh P, Benjak A, Schuenemann VJ, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proc Natl Acad Sci 2015;112(14):4459–64.
7. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 2008;4(1):1–13.

Chapter 4.4 – Phylogenomics and antimicrobial resistance of the leprosy bacillus *Mycobacterium leprae*

Andrej Benjak[1], Charlotte Avanzi[1], Pushpendra Singh[1,2], Chloé Loiseau[1,3], Selfu Girma[4], Philippe Busso[1], Amanda N Brum Fontes[5], Yuji Miyamoto[7], Masako Namisato[9], Kidist Bobosha[4], Claudio G Salgado[10], Moisés B da Silva[10], Raquel C Bouth[10], Marco A C Frade[11], Fred Bernardes Filho[11], Josafá G Barreto[12], José A C Nery[13], Samira Buhrer-Sekula[14], Andréanne Lupien[1], Abdul R Al-Samie[15], Yasin Al-Qubati[15], Abdul S Alkubati[15], Gisela Bretzel[16], Lucio Vera-Cabrera[17], Fatoumata Sakho[18], Christian R Johnson[19], Mamoudou Kodio[20], Abdoulaye Fomba[20], Samba O Sow[20], Moussa Gado[21], Ousmane Konaté[21], Mariane M A Stefani[14], Gerson O Penna[22], Philip N Suffys[5], Euzenir N Sarno[13], Milton O Moraes[13], Patricia S Rosa[23], Ida M Foschiani Baptista[23], John S Spencer[24], Abraham Aseffa[4], Masanori Matsuoka[7], Masanori Kai[7], & Stewart T Cole[1,*]

[1]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.
[2]Department of Microbiology and Biotechnology Centre, Maharaja Sayajirao University of Baroda, Vadodara, India. (Present address).
[3]Department of Medical Parasitology and Infection Biology, Swiss Tropical and Public Health Institute, Basel, Switzerland. (Present address).
[4]Armauer Hansen Research Institute, Addis Ababa, Ethiopia.
[5]Laboratory of Molecular Biology Applied to Mycobacteria, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil.
[7]Leprosy Research Center, National Institute of Infectious Diseases, Higashimurayama, Tokyo, Japan.
[9]AUEN Polyclinic, Nakatomi, Tokorozawa, Saitama Prefecture, Japan.
[10]Laboratório de Dermato-Imunologia, Instituto de Ciências Biológicas, Universidade Federal do Pará, Marituba, Pará, Brazil.
[11]Dermatology Division, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil.
[12]Spatial Epidemiology Laboratory, Federal University of Pará, Castanhal, Pará, Brazil.
[13]Leprosy Laboratory, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil
[14]Tropical Pathology and Public Health Institute, Federal University of Goiás, Goiânia, Brazil.
[15]Ministry of Health and Population, c/o National Leprosy Elimination Programme, Taiz, Republic of Yemen.
[16]Department of Infectious Diseases and Tropical Medicine, Medical Center of the Ludwig-Maximilians-University, Munich, Germany.
[17]Laboratorio Interdisciplinario de Investigación Dermatológica, Servicio de Dermatología, Hospital Universitario, Monterrey, N.L., Mexico.
[18]Programme National Lèpre de Guinée, Conakry, Guinea.
[19]Centre Interfacultaire de Formation et de Recherche en Environnement pour le Développement Durable, University of Abomey-Calavi, Cotonou, Benin.
[20]Centre National d'Appui à la Lutte Contre la Maladie, Bamako, Mali.
[21]Programme National de Lutte contre la Lèpre, Niamey, Niger.
[22]Tropical Medicine Centre, University of Brasília, Fiocruz, Brasília, Brazil.
[23]Lauro Souza Lima Institute Bauru, São Paulo, Brazil.
[24]Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado, USA.

2018. Nat Com. accepted

Contributions: study design, sample processing, DNA extraction, library preparation, array capture, interpretation of results, manuscript preparation

Abstract: Leprosy is a chronic human disease caused by the yet-uncultured pathogen *Mycobacterium leprae*. Although readily curable with multidrug therapy (MDT), over 200,000 new cases are still reported annually. Here, we obtain *M. leprae* genome sequences from DNA extracted directly from patients' skin biopsies using a customized protocol. Comparative and phylogenetic analysis of 154 genomes from 25 countries provides insight into evolution and antimicrobial resistance, uncovering lineages and phylogeographic trends, with the most ancestral strains linked to the Far East. In addition to known MADT-resistance mutations, we detect other mutations associated with antibiotic resistance, and retrace a potential stepwise emergence of extensive drug resistance in the pre-MDT era. Some of the previously undescribed mutations occur in genes that are apparently subject to positive selection, and two of these (*ribD*, *fadD9*) are restricted to drug-resistant strains. Finally, nonsense mutations in the *nth* excision repair gene are associated with greater sequence diversity and drug resistance

Mycobacterium leprae is the main causative agent of leprosy, a disease that affects the skin, nerves, and mucosa of the upper respiratory tract in humans[1]. A second, distantly related leprosy bacillus, *Mycobacterium lepromatosis*, was recently discovered in humans and red squirrels (*Sciurus vulgaris*)[2]. Leprosy is curable with multidrug therapy (MDT), but remains a public health problem in South America, Africa, South and Southeast Asia, and Micronesia, where over 200,000 new leprosy cases are reported each year[3]. MDT, comprising rifampicin, dapsone, and clofazimine, has been used intensively since the 1980s and a few second-line drugs, ofloxacin, minocycline, and clarithromycin, are sometimes employed as therapeutic agents[4]. The emergence of drug-resistant (DR) and multidrug-resistant (MDR) *M. leprae* is increasingly reported[5–12]. For dapsone, rifampicin and ofloxacin, the resistance mechanism has been attributed to missense mutations in the drug resistance determining regions (DRDR) of the *folP1*, *rpoB*, and *gyrA* genes, respectively.

M. leprae is an obligate intracellular pathogen that has never been cultured axenically but can infect wild or experimental animals. The nine-banded armadillo (*Dasypus novemcinctus*) or the mouse footpad (MFP) can be used to produce bacilli, but both methods are cumbersome and time-consuming[13]. The genome of *M. leprae* is the smallest among mycobacteria (3.3 Mb) with 1614 genes encoding proteins and a remarkable 1300 pseudogenes[14]. Such reductive evolution is a hallmark of bacteria that have changed their lifestyle from free-living to strictly hostassociated[15]. Due to its 14-day generation time and the absence of horizontal gene transfer, the genome of *M. leprae* is highly conserved, with <300 single-nucleotide polymorphisms (SNPs) observed between distantly related strains, and only a few SNPs between close relatives[5,16–18]. Four SNP types (branches 1–4) and 16 SNP subtypes (A–P) were defined by surveying 78 informative SNPs and six single-base insertion/deletions (InDels)[16,19]. Genotyping a large panel of *M. leprae* strains revealed strong geographical associations and suggested possible routes of dissemination of leprosy[16] whereas, a recent phylogenetic analysis of 16 whole-genome sequences of modern and ancient *M. leprae* strains, implicated the 3K subtype (branch 0) as the most ancestral[17]. Leprosy seems to have appeared during the Iron Age (1200–600 BC) and the date of the most recent common ancestor of *M. leprae* was estimated to be from 2543 BC to 36 AD, based on whole-genome sequence analysis[17]. Similarly, the earliest accepted written record of leprosy is from 600 BC[20], and the earliest osteological evidence dates from around 300 BC[21–25]. The oldest genomic evidence of

leprosy is for samples from 80 to 240 AD in Central Asia[26]. In this study, we develop and apply methods to isolate and purify *M. leprae* DNA that enable whole genome sequences to be obtained directly from human biopsy material, thus removing the necessity for passage through animals. This approach was successfully used to generate 120 new *M. leprae* genome sequences from drug-susceptible and DR strains from around the world, thereby enabling detailed phylogenetic and phylogeographic comparisons to be performed, new mutations associated with antimicrobial resistance to be detected, and the likely origin of leprosy to be proposed.

Results

Isolating *M. leprae* DNA from human skin biopsies. Genome sequencing has become routine practice in microbiology[27], especially for micro-organisms that can be readily isolated, which is not the case of the leprosy bacillus. For decades, the sole source of *M. leprae* DNA suitable for genomics was from bacteria isolated 12 months after infection of armadillos or mice. Recently, we have developed and optimized methods that enable *M. leprae* DNA to be extracted directly from fresh or formalin-fixed skin biopsies from leprosy patients[28]. These methods include enrichment of *M. leprae* DNA by array capture[17] but this is less practical for large population-based investigations. The DNA extraction method used in this study was applied directly to punch biopsies from clinically well-characterized patients of known bacillary index (BI) and exploits the fact that *M. leprae* resides intracellularly. Host cells are first disrupted and their DNA degraded, leaving the bacilli intact. The bacilli are then lysed and their DNA extracted and used for library preparation. This approach was applied to 106 biopsies whose BI ranged from 0 (no bacilli visible) to 6 (>1000 bacilli per microscopic field) thereby enabling a relationship between BI and sequencing efficiency to be established (Fig. 1). As expected, there was a direct correlation between genome coverage and the BI but, surprisingly, successful coverage could even be achieved with some specimens whose BI was as low as 1+.

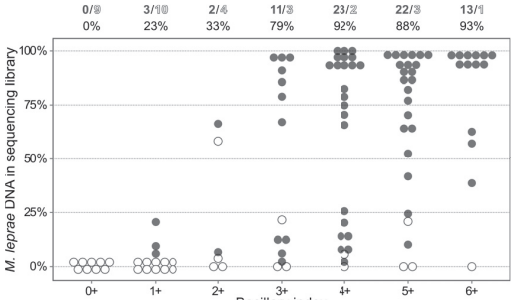


Figure 1: Correlation between bacillary index and successful sequencing. The content of *M. leprae* DNA in sequencing libraries derived from human skin biopsies was determined and found to be proportional to the bacillary index (not available for all samples). Empty circles are samples that were not included in the study due to insufficient genome coverage. Sample count and sequencing success rates are given at the top of each category

Genome analysis of patient and animal cohort. We analyzed a total of 154 *M. leprae* genomes from 25 countries (Fig. 2, Supplementary Data 1), of which 120 were newly sequenced and 34 were previously published (Supplementary Data 1). The cohort comprised 147 human samples, 6 from red squirrels and 1 from an armadillo that were all naturally infected. Genome sequences were obtained directly from 109 human samples, 30 from bacilli passaged in mice, and 8 from armadillos. Thirty of these strains were from patients who had relapsed or not responded to MDT the remainder (124) were from supposedly drug-susceptible strains (87 were from confirmed primary cases, while disease history was unknown for the others).

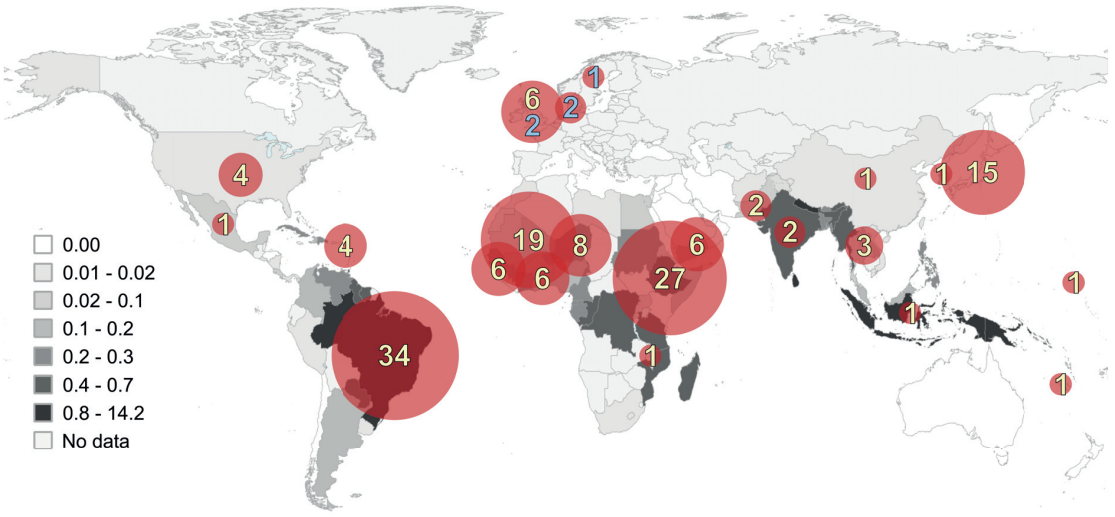


Figure 2: Geographic distribution of the *M. leprae* samples used in this study. World map shows the number of registered cases of leprosy per 10,000 population (prevalence rates) in 2015 as reported by the World Health Organization (http://apps.who.int/neglected_diseases/ntddata/leprosy/leprosy. html). Blue numbers indicate ancient *M. leprae* strains

A total of 3053 SNPs and 219 InDels (excluding tandem repeats) was found (Supplementary Data 2). The average SNP difference among the 154 genomes was 114. We found a total of 988 non-synonymous alleles (0.62 per protein-coding gene, or 0.61 per kb of protein-coding genes) and 530 synonymous SNPs (0.33 per protein-coding gene or 0.33 per kb of protein-coding genes), and 1763 mutations in intergenic regions andpseudogenes (1.07 mutations per kb of intergenic regions and pseudogenes). The SNP density for each gene is given in Supplementary Data 2. Of the 219 InDels, 58 (27%) were in protein-coding genes.

Phylogeny of *M. leprae*. Phylogenetic analysis using both maximum parsimony (MP) and Bayesian inference resulted in consistent tree topologies and revealed distinct lineages and sublineages of *M. leprae* (Fig. 3). Strains belonging to the same SNP subtypes[16] clustered within single branches, with the exception of SNP subtype 3K, which is represented by a newly discovered ancestral lineage, termed here 3K-1, and the ancestral lineage referred to earlier as branch 0[17] and termed here 3K-0 (Fig. 3a). All strains from the two most ancestral lineages, 3K-0 and 3K-1, originated from Japan (8), China (1), Korea (1), the Marshall Islands (1), and New Caledonia (1), in agreement with earlier genotyping studies of hundreds of *M. leprae* strains which confirmed the predominance of the 3K genotype in East Asia, notably in Japan, China, and Korea[16,29–31]. *M. leprae* in East Africa showed higher diversity with subtypes 2E, 2F, and 2H representing distinct lineages (Fig. 3a). The geographic distribution of those lineages corroborates earlier studies reporting the presence of SNP type 2 in Medieval Europe, the Middle East and East Africa[16]. Two Ethiopian isolates, belonging to the 2F subtype, clustered closely with medieval European strains dating from the 11th to 12th century (Fig. 3), which supports the hypothesis that the ancient Greek and Roman routes[32] connecting Europe, the Middle East, East Africa, and South Asia[16,33–35] contributed to the dissemination of SNP type 2*M. leprae*. West Africa, on the other hand, harbors exclusively SNP type 4, suggesting that overland migration between East and West Africa was limited. SNP subtypes 4N, 4O, and 4P, albeit sharing the same ancestor, do not form a monophyletic clade as previously hypothesized[16]. Rather, the 4O and 4P subtypes cluster together in a branch

distinct from 4N (Fig. 3). Brazil, as expected, contains a great diversity of several *M. leprae* lineages, with the SNP type 4 and SNP subtype 3I being the most prevalent[36]. The 3I genotype was common in medieval Europe[17,21,37,38], and is still present in red squirrels in the United Kingdom[2]. The modern Brazilian strain Br2016-45 branched between two medieval strains from Europe (Fig. 3), making it the most ancestral contemporary 3I strain in the Americas to date. The broad diversity of 3I genotypes from Brazil probably derives from multiple introductions from Europe. On the other hand, the strains circulating in the Southern USA and associated with zoonosisfrom the nine-banded armadillo, I-30, NHDP-55 and NHDP-63[18], originated much more recently (Fig. 3b), in agreement with the rapid expansion and spread of the armadillo population since its introduction to this region about 150 years ago[39]. Good representation of most *M. leprae* lineages enabled identification of lineage-specific markers. A set of 235 SNPs and 25 InDels were specific to single lineages or groups of related lineages (Supplementary Data 2), of which 73 non-synonymous SNPs and 5 InDels were within protein-coding genes. These new lineage-specific markers can be used for future genotyping schemes.

Table 1 Mutations in genes associated with drug resistance.

Sample	*folP1* ML0224	*rpoB* ML1891c	*gyrA* ML0006	*gyrB* ML0005	*fadD9* ML0484c	*ribD* ML1340	*ethA* ML0065	*pks4* ML1229	*nth* ML2301c
2188-2007	.	.	.	.	.	G62D[1]	.	.	.
2188-2014	.	.	.	.	.	G62D	.	.	.
85054	**P55L**	**S456L** G52E	.	.	W878*	D256N	.	.	R197*
2DDS	**P55R**	.	.	.	L396P	S58R	.	.	.
LRC-1A	.	.	I851T	V214G	.	.	.	.	.
Airaku-3	**T53I**	.		.	N304fs	.	.	.	.
Amami	**P55L**	.	.	.	.	R236C	.	.	L145fs
ARLP_08	.	H200Y	.	.	.	.	.	.	.
ARLP_10	.	.	.	.	Y927D	G61C	.	.	.
ARLP_30	.	.	G1115R	.	.	.	.	.	.
ARLP_52	.	.	G1115R	.	.	.	.	.	.
Bn8-46	**P55R**	.	.	.	.	.	.	.	.
Bn8-52	**P55L** N145H	.	.	.	W1108fs	.	.	.	.
Br14-1	**P55R**	**S456M**	.	.	G148fs	.	.	.	.
Br14-2	**P55R**	**S456M**	.	.	G148fs	.	.	.	.
Br14-3	**P55L**	**H451Y** G448D T433I	**A91V**	.	Q107*	G94D	.	.	E173*
Br14-4	**P55R**	**S456L** Y171N	.	.	L396P	I56T	.	.	.
Br14-5	**P55R**	**S456L**	.	.	A919E	C222W	.	.	.
Br2016-15	**P55L**	.	V731I	T503I	G796S	A63T	.	M14I	N142fs
Br2016-16	.	.	.	.	L998fs	K267fs	.	.	.
Br2016-18	.	.	S307L	.	.	.	.	.	.
Br2016-21	.	.	.	.	.	.	.	I932fs	.
Kusatsu-6	**P55L**	T433I **D441Y**	.	.	.	.	K477fs	.	.
MI10-93	**P55R**	.	.	.	R73fs	S58R	.	.	.
MI10-98	.	.	.	.	.	A10fs	.	.	.
MI2-10	**T53R**	.	.	.	.	.	.	.	.
MI6-50	**T53R**	.	.	.	.	.	.	.	.
MI6-55	**T53R**	.	.	.	.	.	.	.	.
Ng14-35	**P55R**	.	S307L	.	A594T	D34del	.	.	.
S15	**T53I**	G432S **H451**D	.	.	D466N	Q117*	D63N D323N	T334I	G146fs
S9	**T53I**	R791Q	.	.	.	S58N	.	.	.
Thai-237	.	.	I851T	.	.	.	.	.	.
US57	.	.	G362E	.	.	.	.	.	.
Zensho-2	**P55L**	.	.	.	Y562fs	G94D	.	.	.
Zensho-4	**T53I**	P51S **S456L**	**A91V**	D464N	R314C	D77N	A25T	Q1719*	L163fs
Zensho-5	**P55L**	**S456L** P583L	.	.	.	G204C	G390A	.	N142fs
Zensho-9	**P55L**	**H451Y** G681S	.	.	A973T	P150L	P383L	.	E122*

In bold, substitutions or residues known to confer drug resistance in *M. leprae*; *, premature stop codon; fs, frameshift; dot (.), wild-type. RpoB numbering is based on *M. leprae*, *E. coli* numbering in brackets: 51(126), 52(127), 171(246), 200(275), 432(507), 433(508), 441(516), 448(523), 451(526), 456(531), 681(756), 791(866). [1]Coverage below the threshold, both "2188" isolates come from the same patient but after an interval of 7 years[28].

Dating analysis. Dating analysis of *M. leprae* (Fig. 3b and Supplementary Figure 1) was done using BEAST v2.4.4[40] and the results were very similar to those obtained from a ten-fold smaller number of contemporary isolates[17]. The most recent common ancestor (TMRCA) of all *M. leprae* strains was estimated to be 3699 years old (95% Highest Posterior Density (HPD) 2731–4838 ya) and the substitution rate was 7.8×10^{-9} per site per year. Overlapping results were obtained when using different models (Supplementary Note and Supplementary Table 1), indicating that the data set was robust and sufficiently informative.

Table 2 Highly polymorphic genes and genomic regions of *M. leprae*.

Gene or region	Description	Non-synonymous mutations (multi-allele loci)	Synonymous mutations	Homo-plasy
ML0411	Serine-rich antigen	28 (4)	1	4
ML1040c	Putative ATP-dependent helicase	18 (1)	0	0
fadD9	Probable fatty-acid-CoA ligase	17	1	1
ML1750c	Putative nucleotide cyclase	17	1	0
ML1512c	Putative ribonuclease J	17	0	0
ribD	Bifunctional enzyme riboflavin biosynthesis protein	14	0	1
rpoB	DNA-directed RNA polymerase (beta chain)	12 (1)	0	3
ML1753c	Probable transcriptional regulatory protein	9	0	0
gyrA	DNA gyrase (subunit A)	8	8	1
ML1300	Conserved hypothetical protein	8	0	0
nth	Endonuclease III	7	0	1
ctpC	Metal cation-transporting P-type ATPase C	7	2	0
rpoC	DNA-directed RNA polymerase (beta chain)	7	5	0
ML2687c	Probable conserved transmembrane protein	7	0	0
ML1052c	Conserved hypothetical protein	7	1	0
ML0009	Hypothetical protein	6	1	0
ML0283	Cation-efflux transporter component	6	1	0
ML2700	Conserved transmembrane protein	6	7	0
mfd	Transcription-repair coupling factor	6	1	0
ppsC	Phenolphthiocerol and DIM synthesis	6	1	0
ethA	Activates the pro-drug ethionamide	5	0	0
trpE	Biosynthesis of tryptophan (at the first step)	5	1	0
pknB	Transmembrane serine/threonine-protein kinase	5	3	0
fas	Fatty acid synthase	5	1	0
esxA	Early secretory antigenic target	2	0	1
lsr2	Dominant T-cell antigen and stimulates lymphoproliferation.	4	1	3
folP1	Dihydropteroate synthase	3 (2)	1	2
ML1752c	Conserved hypothetical protein	2	1	1
ppsA	Phenolphthiocerol and DIM synthesis	4	3	1
ML0803	Two-component sensor kinase	1 (1)	0	0
ML0237	Conserved hypothetical protein (pseudogene)	NA	1	1
ML0010c-ppiA	Intergenic region	NA	3	1

A striking observation is the relative youth of the SNP type 1 lineage and its association with South Asia (Fig. 3b). Earlier studies revealed a predominance of SNP subtype 1D in India and Nepal, followed by 1C, 1A[16], and 2E, 2G, and 2H[33–35]. SNP type 1 predominates in Thailand[41], Bangladesh, Indonesia, and Philippines[16]. The current phylogeography of *M. leprae* implies that humans brought leprosy to South Asia from other parts of the continent.

Hypermutated *M. leprae* strains. Eight *M. leprae* strains (85054, S15, Amami, Zensho-4, Zensho-5, Zensho-9, Br14-3, and Br201615), belonging to five different SNP subtypes, had unusually long branches in the MP tree (Fig. 3a) because they contained on average 92 more SNPs than the other strains but approximately the same number of InDels. Comparative analysis revealed one unique feature linking the observed "hypermutated" strains, namely deleterious mutations in the endonuclease III gene *nth* (*ML2301*) due to frameshifts and premature stop codons (Table 1).

Drug resistance. DR-associated SNPs were detected in the DRDR in 24 strains for *folP1*, 11 strains for *rpoB*, and 2 strains for *gyrA* (in bold in Table 1). Previously described mutations were identified in *folP1* at codons 53 (n = 7) and 55 (n = 17), except in one isolate (Bn8-52), which had mutations at codons 55 and 145. Eleven strains had known mutations that confer rifampicin resistance in their *rpoB*-DRDR, while two strains (Kutatsu-6 and S15) harbor one additional mutation, and one (Br14-3) has two additional mutations in the DRDR (Table 1). One of

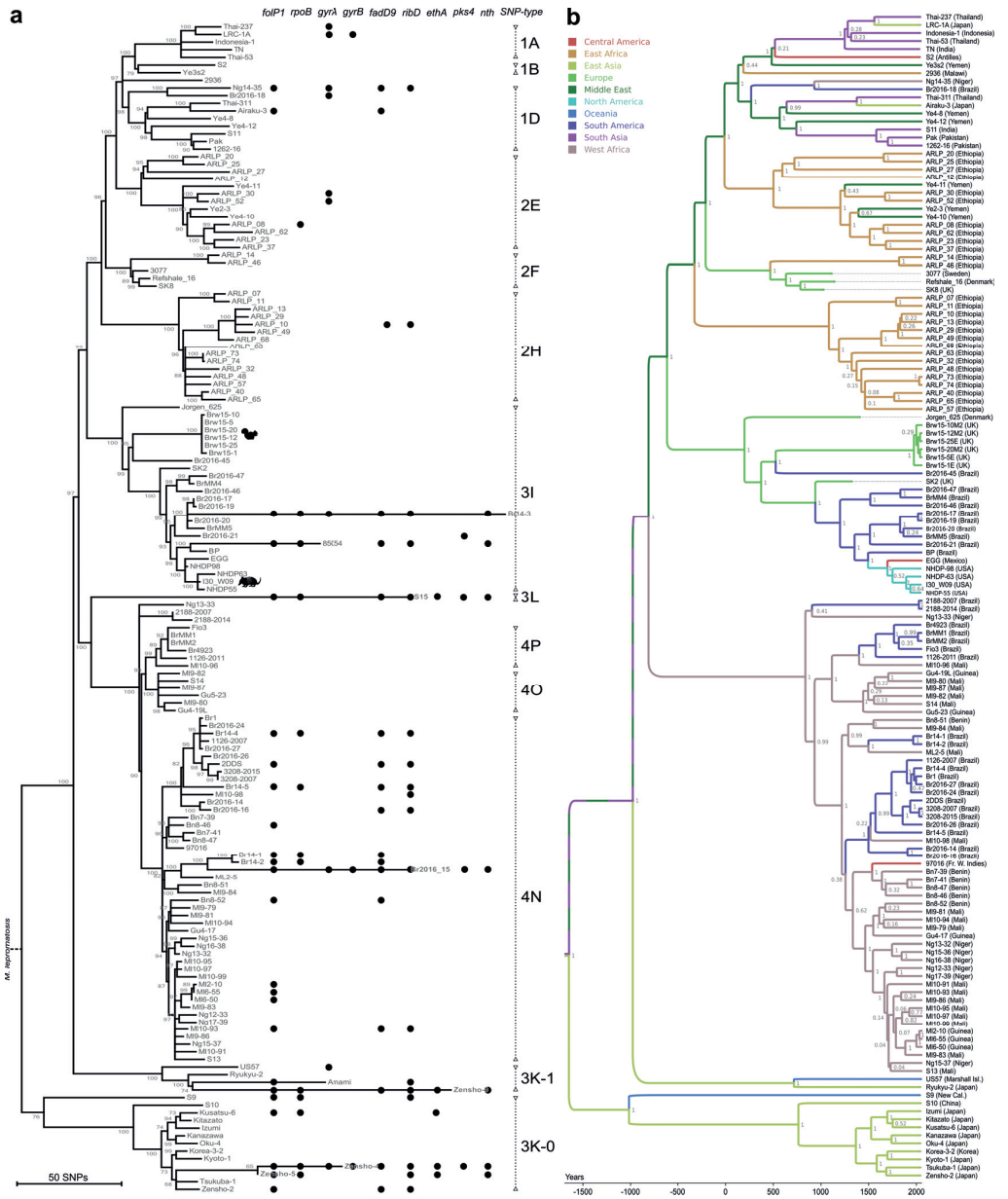


Figure 3: Phylogeny of *M. leprae*. a Maximum parsimony tree of 154 genomes of *M. leprae*. The tree is drawn to scale, with branch lengths representing number of substitutions. *M. lepromatosis* was used as outgroup. Bootstrap values (500 replicates) are shown next to the branches. Dots indicate protein-changing mutations in the corresponding gene as given in Table 1. b Bayesian phylogenetic tree of 146 genomes of *M. leprae* calculated with BEAST 2.4.4. Hypermutated samples with mutations in the *nth* gene were excluded from the analysis. The tree is drawn to scale, with branch lengths representing years of age. Samples were binned according to geographic origin as given in the legend. Posterior probabilities for each node are shown in gray. Location probabilities of nodes were inferred by the Discrete Phylogeny model

these additional mutations (G432S) does not confer rifampicin resistance to recombinant *Mycobacterium smegmatis*[42] whereas no information is available for the remaining three (T433I, G448D, and T508I) except that the G448A substitution does confer rifampicin resistance in *M. tuberculosis*[43]. Also, 5 of the 11 rifampicin-resistant strains had additional missense mutations in *rpoB* (85054, Br14-4, Zensho-4, Zensho-5, and Zensho-9) while 2 strains (ARLP_08 and S9) presented non-synonymous SNPs outside the DRDR (Fig. 4). Compensatory mutations in *rpoA* and *rpoC*, encoding the alpha and beta-prime subunits of RNA polymerase, can occur in rifampicin-resistant *M. tuberculosis*[44]. We found one non-synonymous SNP in *rpoA*, substitution T187P in the rifampicin-resistant strain Br14-5, and seven nonsynonymous SNPs in *rpoC* (Supplementary Data 2), of which two occurred in the drug-resistant strains S15 (A258T) and Zensho-4 (H1133Y).

Two strains had known quinolone resistance mutations in the DRDR of *gyrA* and six harbored different single mutations elsewhere in the gene. Three isolates had a missense mutation in *gyrB*, including two within the DRDR (Table 1). Five strains harbor deleterious mutations in the *ethA* gene, encoding a monooxygenase that activates thioamide prodrugs in *M. tuberculosis*[45,46]. Interestingly, in addition to *ethA* and *nth*, three genes (*fadD9*, *ribD*, *pks4*) were mutated almost exclusively in MDR strains occurring 18, 19, and 4 times, respectively (Table 1).

Retracing the emergence of drug resistance in leprosy patients. Prior to the introduction of MDT in the 1980s, patients were treated with dapsone or other antimicrobials as

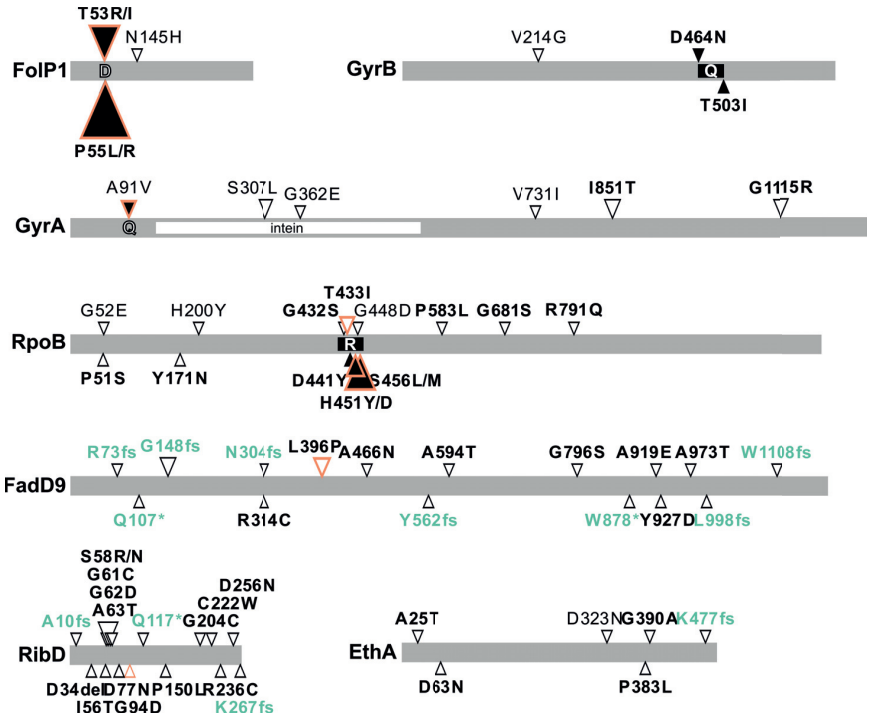


Figure 4: Mutations of *M. leprae* genes associated with antimicrobial resistance. Triangles point to the location of the mutation in the protein. Black triangles indicate known resistance-conferring mutations identified in this study that are situated in the drug resistance determining regions (DRDR): D dapsone, Q quinolone, R rifampicin. Orange border means the mutation was found to be homoplasic. Triangle size reflects the number of isolates from this study harboring the mutation, ranging from 1 to 17. Frameshifts and premature stop codons are in turquoise. Substitutions predicted to have an impact on the biological function of the protein[75] are in bold. Proteins are drawn to scale

monotherapies of varying duration[7,8]. Since genomics uncovered new mutations that are associated with antimicrobial resistance in other bacteria, such as those in *ethA* and *gyrB*, this prompted us to try and retrieve the clinical records of six patients whose strains displayed resistance to three or more drugs (dapsone, rifampicin, quinolones, and thioamides). Four of these extensively drug-resistant (XDR) strains were from multibacillary patients in Japan who had received a succession of monotherapies in the pre-MDT era and our genome analysis enabled the chronology of resistance emergence to be retraced. This is illustrated in Fig. 5, and sadly exemplified by the strain from patient Zensho-4 who was diagnosed in 1963 and first treated with protionamide followed by thiambutosin, both of which show cross-resistance and likely require activation by the EthA mono-oxidase that acquired the A25T missense mutation[47]; then treatment began with dapsone leading to emergence of the T53I mutation in *folP1*, followed by rifapentine that selected the S456L mutation in *rpoB*, and continued with ofloxacin, to which resistance arose from the A91V mutation in *gyrA* and D464N in *gyrB*. Molecular drug susceptibility testing was performed in 1998 and the patient finally cured by a regimen comprising clofazimine, minocycline, chloramphenicol and levofloxacin/sparfloxacin. The fifth XDR strain was from a newly diagnosed Brazilian case (Br2016-15) with no history of treatment for leprosy, confirming the ongoing transmission of primary antimicrobial resistance, while details of the sixth case could not be recovered.

Genes under positive selection. We also identified genes containing an unusually high number of polymorphisms, multiple alleles, and homoplasies (Supplementary Figure 2), which could be indicative of positive selection[48]. Strikingly, the distribution of these polymorphic sites around the genome was not random as they were often clustered, especially proximal to either side of the origin of replication (Supplementary Figure 3). Protein-changing mutations were found in 540 genes, with an average of 1.77 mutations per gene (STD 2.12). Table 2 contains a ranking of genes with at least five non-synonymous mutations or regions with one or more homoplasy (excluding VNTRs). The most polymorphic gene by far was *ML0411*[49] encoding the serine-rich antigen, a member of the immunogenic, surface-exposed PPE protein family. Two other known T-cell antigens whose genes display variability are Lsr2 and EsxA (Table 2). Other than *nth*, three other polymorphic genes (*ML1040c*, *ML1750c*, and *ML1512c*) code for proteins that appear to function in nucleic acid or cyclic nucleotide metabolism (Table 2).

Discussion

Here we have optimized and applied highly sensitive procedures to extract *M. leprae* DNA directly from human skin biopsies that is suitable for whole-genome sequencing. The resultant genome sequences were analyzed phylogenetically and used to retrace the origin of the leprosy bacillus, and to identify polymorphisms that had been positively selected during evolution. Such polymorphisms might reflect pressure from the human immune system, from MDT or other forces.

It is striking that the ancestral lineages of *M. leprae* predominate in East Asia, although we should keep in mind that Central Asia has been understudied, so it would be interesting to sequence more samples spanning the East–West axis of Asia, including the Middle East, where the 3K genotype is also present[16]. Nevertheless, given the current data on the distribution of the 3K subtype we can deduce that the ancestor of *M. leprae* originated within Eurasia, probably in the Far East.

Endonuclease III (Nth) and the formamidopyrimidine and endonuclease VIII family (Fpg/Nei) of DNA glycosylases are

central to the base excision repair pathway in bacteria[50]. Mycobacterial genomes usually contain a single *nth* and two *fpg/nei* genes but *M. leprae* has lost both *fpg/nei* orthologues and retained the *nth* gene. Nth, Fpg, and Nei may have overlapping functions and, in enteric bacteria, mutator phenotypes were observed when *nth* was inactivated in combination with the *fpg* and *nei* genes[51– 53]. In *M. smegmatis*, deletion of *nth* and both the *nei* homologs resulted in elevated spontaneous mutation frequencies and increased sensitivity to oxidative stress[54]. Therefore, in the absence of Nei, inactivation of *nth* in *M. leprae* should lead to increased sequence variability, which is consistent with our results.

Strikingly, all *nth* mutants were also drug-resistant so Nth loss likely favors emergence of drug resistance, and *nth* mutations might serve as a surrogate marker for potential drug resistance and treatment failure. A link between a higher mutation rate and drug resistance was observed in strains of *M. tuberculosis* (which has *nth* and two *fpg/nei* genes) belonging to lineage 2, but the molecular basis for this is unknown[55]. For a pathogen with an extremely reduced genome such as *M. leprae*, a hypermutator phenotype could be detrimental and ultimately lethal.

Drug resistance is alarming for leprosy control. There is growing evidence for primary quinolone resistance in strains of *M. leprae* from patients who have never been treated with quinolones for leprosy but may have received this drug for other infections[56]. Five new GyrA mutations were identified in this study, but their effect on FQ resistance remains to be determined. Since two of them arose independently in the GyrA intein, which is removed by protein splicing, they may not impact quinolone activity (Fig. 4). Three non-synonymous mutations were found in *gyrB* (Table 2) and experimental evidence exists for two of them conferring quinolone resistance in *in vit*ro assays or in *M. tuberculosis*[57,58]. To our knowledge, this is the first report of *M. leprae* clinical isolates harboring mutations in *gyrB*. Thus, despite their apparent rarity, mutations in *gyrB* should be systematically assessed in drug resistance screening. A range of known and new mutations was detected in the DRDR and elsewhere in *rpoB* (Fig. 4; Table 1). Some of these might have a compensatory role in restoring fitness, that is known to be reduced to various degrees in rifampicin-resistant mutants of *M. tuberculosis*[59,60]. Similarly, compensatory mutations in *rpoA* and *rpoC* can occur in rifampicin-resistant *M. tuberculosis*[44]. The *rpoA* substitution T187P in the rifampicin-resistant *M. leprae* strain Br14-5 was shown to be compensatory in *M. tuberculosis*[44].

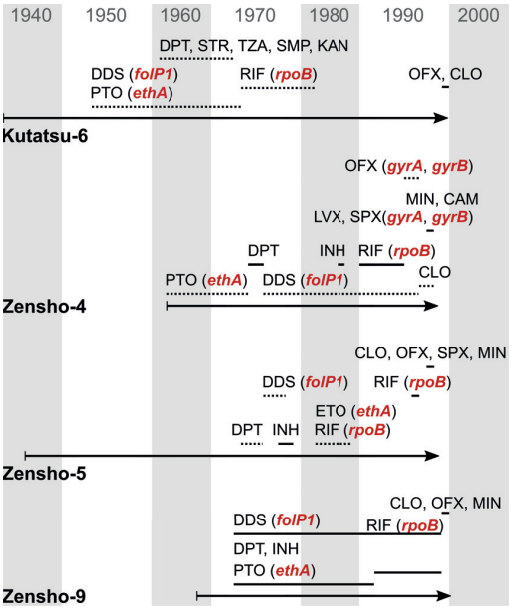


Figure 5: Timeline of the leprosy treatment and emergence of drug resistance in the XDR strains. Mutated genes conferring resistance to the corresponding drugs are shown in red. Arrows span from the onset of disease to the end of treatment. Horizontal lines show the period when a drug was given. Dotted lines mean irregular treatment. CAM chloramphenicol, CLO clofazimine, DDS dapsone, DPT thiambutosine (diphenylthiourea), ETO ethionamide, INH isoniazid, KAN kanamycin, LVX levofloxacin, MIN minocycline, OFX ofloxacin, PTO protionamide, RIF rifampicin, SMP sulfamethoxypyridazine, SPX sparfloxacin, STR streptomycin, TZA thiozamin

Rifampicin-resistant isolates of *M. tuberculosis* harbor more mutations in *rpoC*

compared to rifampicinsusceptible isolates[44,61]. In our case, we observed no clear correlation between rifampicin resistance and mutations in *rpoC*, which occurred in two resistant and two wild-type strains.

Arguably the most intriguing finding of the present investigation was the remarkably high frequency of mutations in the *fadD9* and *ribD* genes and in 19/23 cases these occur in strains that have at least one mutation that is associated with resistance to a leprosy drug (Fig. 4; Table 1). Functional information for *fadD9* is scarce and the mutations found are predicted to either abolish protein production (8/16) or to cause detrimental amino acid changes (Fig. 4). In the case of *ribD*, 14 different missense mutations were found in a group of 17 variant alleles, indicating that this is likely an essential function. From studies with *M. tuberculosis* it is known that *ribD* encodes an alternative dihydrofolate reductase, with relatively low activity compared to that conferred by the bona fide dihydrofolate reductase gene, *dfrA*[62]. In clinical isolates of *M. tuberculosis*, a promoter mutation causes overexpression of *ribD* that is associated with resistance to the old drug, *para*-amino salicylic acid (PAS), and to certain DHFR inhibitors[63]. This suggests that the mutations detected in the *M. leprae ribD* gene may also confer resistance to PAS and support for this is provided by the fact that vadrine (2-pyridyl-(4)-1,3,4oxydiazolone-(5)-*p*-aminosalicylate) was used as a drug to treat leprosy before dapsone became widely available[64]. It is thus possible that the *ribD* mutations we report here arose nearly 60 years ago following treatment with vadrine or another PAS derivative. Our discovery of these mutations and those in *fadD9* should encourage further experimentation in order to establish their true role and contribution to antimicrobial resistance, especially to clofazimine.

Methods

Sample collection. Samples were taken from leprosy patients as punch biopsies of skin (preserved in 70% ethanol or formalin-fixed and paraffin-embedded (FFPE)), which is standard diagnostic procedure for leprosy, or from mouse foot-pads. Details about the samples used in this study are given in Supplementary Data 1 and below.

Origin of S15: Strain S15 corresponds to strain 92041[65], which was isolated from a lepromatous leprosy patient originally from Martinique. The origin of S15 was erroneously attributed to New Caledonia in Monot et al.[16], and the error was subsequently propagated in several publications[2,17,66,67].

DNA extraction and library preparation. DNA was extracted from 101 human skin biopsies with known BI using a customized in-house protocol combining host tissue digestion and the QIAmp microbiome kit for host DNA depletion, strong bacterial cell lysis and silica-based purification. Punch biopsies (6 mm) in 70% ethanol were first rehydrated in Hank's balanced solution prior to mincing with scissors. Cells were detached from the tissue by 30 min incubation at 37 °C with a mixture of 0.5 U of collagenase and dispase, followed by incubation at 56 °C with 10 mg/ml of trypsin until complete digestion. Free cells were then suspended in 1 ml of phosphate-buffered saline (PBS) and DNA was extracted using the QIAmp DNA microbiome extraction kit according to the manufacturer's recommendations. Each run of extraction included a batch of five to nine samples and one blank control (500 µl of Hank's balanced solution). The presence of *M. leprae* was assessed by PCR using RLEP primers[2] prior to library preparation. Libraries were prepared from 50 µl of extracted DNA using the Kapa Hyperprep kit as described previously[2,5]. DNA from FFPE samples was extracted using the truXTRACTM FFPE DNA kit (Covaris) as described previously[28]. Libraries prepared from the extracted DNA

were used directly for shotgun sequencing. *M. leprae* DNA extraction quality was assessed from the percentage of *M. leprae* DNA present in the library inferred by alignment to the reference genome sequence, with a minimum threshold set at 1%. This threshold was chosen because it yields an average genome coverage of at least 5× per sample in a multiplexed run of 10 samples on one HiSeq 2500 lane (yielding around 20 million reads per sample and 100 bases per read).

Library enrichment. Libraries with low *M. leprae* content underwent enrichment using whole-genome tiling arrays as described previously[17]. Briefly, Illumina libraries were hybridized onto custom Agilent SureSelect Capture Arrays containing ca. one million DNA probes (60 bp) spanning the entire *M. leprae* genome (tiled every 4 bp), followed by elution and PCR amplification.

Sequencing. Sequencing was performed on Illumina Hi-Seq 2000, Hi-Seq 2500, or Mi-Seq instrument.

Sequence processing. We took precautions in analyzing the data to avoid falsepositive SNP calls. All raw reads were adapterand quality-trimmed with Trimmomatic v0.33[68]. The quality settings were "SLIDINGWINDOW:5:15 MINLEN:40". Paired-end (PE) data were additionally processed with SeqPrep (https://github.com/jstjohn/SeqPrep) to merge overlapping pairs. This increases the accuracy of sequence in the overlapping area, avoids problems in estimating coverage and creates longer reads, which facilitates InDel calling. Duplicate reads were omitted from downstream analyses. This is especially important for libraries with insufficient *M. leprae* DNA fragments, which is not uncommon for low BI samples or samples that are difficult to process, like FFPE samples. In these cases, library enrichment with array-capture, or very deep sequencing often produce a high number of duplicate reads (DNA fragments that were sequenced multiple times, seemingly increasing the overall genome coverage), with each read having dozens or even hundreds of copies. Such reads will amplify possible artefacts and sequence errors, resulting in false SNP calls.

Sequence analysis. Preprocessed reads were mapped onto the *M. leprae* TN reference genome (GenBank AL450380.1) with Bowtie2 v2.2.5[69]. We filtered out all reads with mapping quality below 8 and omitted repetitive regions in the reference sequence. We also omitted ribosomal RNA (rRNA) genes because alignments in these regions tend to be error prone. This is because rRNA genes are highly conserved in bacteria, so sequences from other species could map to the *M. leprae* reference sequence. This usually happens when the content of *M. leprae* DNA in a sequencing library is scarce and is even more pronounced when libraries with low *leprae* content undergo array-capture. However, because lineage-specific mutations were previously observed in the *M. leprae rrs* gene[30], we manually checked the alignments corresponding to the rRNA genes and added the curated results to Supplementary Data 2.

SNP calling was done using VarScan v2.3.9[70]. To avoid false-positive SNP calls the following cutoffs were applied: minimum overall coverage of five nonduplicated reads, minimum of three non-duplicated reads supporting the SNP, mapping quality score >8, base quality score >15, and a SNP frequency above 80%.

InDel calling was done using Platypus v0.8.171 followed by manual curation. Completed genome sequences of *M. leprae* Br4923 and *Mycobacterium lepromatosis* (GenBank JRPY00000000.1) were aligned against the *M. leprae* TN reference using LAST72 using the default parameters for the former and the gamma-centroid option for the latter.

Mixed samples. A large number of missing values, especially in lineage-specific loci,

points to the presence of more than one strain in a sequencing library.

Although not thoroughly tested, in our opinion mixed data sets are mostly due to technical problems or contamination because in some cases we were able to identify the problematic strains. The possible presence of multiple *M. leprae* strains in single skin lesions was not tested in this study, but we expect it to be extremely low. Overall, a few mixed data sets were detected and some were removed from this study, except for samples that we deemed important and describe below. Nevertheless, results were not biased because loci with mixed alleles were treated as missing values.

Zensho-4 seems to contain a fraction of another strain (possibly around 40%) that is closely related to it. Only a few loci had mixed alleles, and these include the A91V substitution in *gyrA* (supported by 62% of reads) and the D464N substitution in *gyrB* (supported by 41% of reads). The latter was attributed to Zensho-4 for simplicity. Similarly, Zensho-5 seems to contain around 30% of Zensho-4. This is the main reason why we could not detect SNPs specific only to Zensho-5 (Fig. 3a), since such SNPs would be "diluted" with wild-type alleles from Zensho-4 and could not pass the SNP "purity" threshold. We included these two samples in this study because they are multi-drug-resistant and belong to the SNPtype 3K-0. Furthermore, mutations in genes conferring drug resistance from this study match with those from earlier reports of these samples, confirming their identity[73].

Thai-311 contains <20% of an unidentified 3K-0 strain that belongs to the Kyoto-1/Zensho-5 cluster of strains (Fig. 3a). SNP calling was not significantly affected. Finally, sample Ye2-3 contained around 25% of an unidentified strain belonging to SNP-type 4. Because we only have few samples from Yemen, we decided to keep Ye2-3 in this study.

Phylogeny and dating analysis. Concatenated SNP alignments were used for the analyses. MP trees were constructed in MEGA6[74] using 500 bootstrap replicates. Sites with missing data were partially deleted (80% coverage cutoff), resulting in 3046 variable sites used for the tree calculation. The Subtree-Pruning-Regrafting algorithm was used as the MP search method. Dating analysis and discrete phylogeography were done using BEAST2 v2.4.4[40]. Details are given in the Supplementary Note.

Data availability. Sequence data are available from the NCBI Sequence Read Archive (SRA) under accession number SRP072827. Accession numbers for all samples used in this study are given in the Supplementary Data 1. Other relevant data supporting the findings of the study are available in this published article and its Supplementary Information files, or from the corresponding author upon request.

References

1. Britton, W. J. & Lockwood, D. N. J. Leprosy. *Lancet* 363, 1209–1219 (2004).

2. Avanzi, C. et al. Red squirrels in the British Isles are infected with leprosy bacilli. *Science* 354, 744–747 (2016).

3. World Health Organization. Global leprosy update 2015: time for action, accountability and inclusion. *Wkly Epidemiol. Rec.* 91, 405–420 (2016).

4. Scollard, D. M. et al. The continuing challenges of leprosy. *Clin. Microbiol. Rev.*19, 338–381 (2006).

5. Avanzi, C. et al. Transmission of drug-resistant leprosy in Guinea-Conakry detected using molecular epidemiological approaches. *Clin. Infect. Dis.* 63, 1482–1484 (2016).

6. Cambau, E., Perani, E., Guillemin, I., Jamet, P. & Ji, B. Multidrug-resistance to dapsone, rifampicin, and ofloxacin in *Mycobacterium leprae*. *Lancet* 349, 103–104 (1997).

7. Maeda, S. et al. Multidrug resistant *Mycobacterium leprae* from patients with leprosy. *Antimicrob. Agents Chemother.* 45, 3635–3639 (2001).

8. Matsuoka, M., Kashiwabara, Y. & Namisato, M. A. *Mycobacterium leprae* isolate resistant to dapsone, rifampin, ofloxacin and sparfloxacin. *Int. J. Lepr. Mycobact. Dis.* 68, 452–455 (2000).

9. Matsuoka, M., Kashiwabara, Y., Liangfen, Z., Goto, M. & Kitajima, S. A second case of multidrug-resistant *Mycobacterium leprae* isolated from a Japanese patient with relapsed lepromatous leprosy. *Int. J. Lepr. Mycobact. Dis.* 71, 240–243 (2003).

10. Matsuoka, M. et al. The frequency of drug resistance mutations in *Mycobacterium leprae* isolates in untreated and relapsed leprosy patients from Myanmar, Indonesia and the Philippines. *Lepr. Rev.* 78, 343–352 (2007).

11. da Silva Rocha, A. et al. Drug and multidrug resistance among *Mycobacterium leprae* isolates from Brazilian relapsed leprosy patients. *J. Clin. Microbiol.* 50, 1912–1917 (2012).

12. You, E.-Y., Kang, T. J., Kim, S.-K., Lee, S.-B. & Chae, G.-T. Mutations in genes related to drug resistance in *Mycobacterium leprae* isolates from leprosy patients in Korea. *J. Infect.* 50, 6–11 (2005).

13. Lahiri, R. & Adams, L. B. *International Textbook of Leprosy* (2016).

14. Cole, S. T. et al. Massive gene decay in the leprosy bacillus. *Nature* 409, 1007–1011 (2001).

15. Moya, A., Peretó, J., Gil, R. & Latorre, A. Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat. Rev. Genet.* 9, 218–229 (2008).

16. Monot, M. et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. *Nat. Genet.* 41, 1282–1289 (2009).

17. Schuenemann, V. J. et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. *Science* 341, 179–183 (2013).

18. Truman, R. W. et al. Probable zoonotic leprosy in the southern United States. *N. Engl. J. Med.* 364, 1626–1633 (2011).

19. Monot, M. et al. On the origin of leprosy. *Science* 308, 1040–1042 (2005).

20. Dharmendra Leprosy in ancient Indian medicine. *Int. J. Lepr.* 15, 424–430 (1947).

21. Taylor, G. M. et al. *Mycobacterium leprae* genotype amplified from an archaeological case of lepromatous leprosy in Central Asia. *J. Archaeol.*

Sci. 36, 2408–2414 (2009).

22. Donoghue, H. D. et al. Co-infection of *Mycobacterium tuberculosis* and *Mycobacterium leprae* in human archaeological samples: a possible explanation for the historical decline of leprosy. *Proc. Biol. Sci.* 272, 389–394 (2005).

23. Blau, S. & Yagodin, V. Osteoarchaeological evidence for leprosy from western Central Asia. *Am. J. Phys. Anthropol.* 126, 150–158 (2005).

24. Tayles, N. & Buckley, H. R. Leprosy and tuberculosis in Iron Age Southeast Asia? *Am. J. Phys. Anthropol.* 125, 239–256 (2004).

25. Mariotti, V., Dutour, O., Belcastro, M. G., Facchini, F. & Brasili, P. Probable early presence of leprosy in Europe in a Celtic skeleton of the 4th–3rd century BC (Casalecchio di Reno, Bologna, Italy). *Int. J. Osteoarchaeol.* 15, 311–325 (2005).

26. Donoghue, H. D. et al. A migration-driven model for the historical spread of leprosy in medieval Eastern and Central Europe. *Infect. Genet. Evol.* 31, 250–256 (2015).

27. Loman, N. J. & Pallen, M. J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* 13, 787–794 (2015).

28. Stefani, M. M. A. et al. Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases. *PLoS Negl. Trop. Dis.* 11, e0005598 (2017).

29. Weng, X. et al Molecular, ethno-spatial epidemiology of leprosy in China: Novel insights for tracing leprosy in endemic and non endemic provinces. *Infect. Genet. Evol.* 14, 361–368 (2013).

30. Yuan, Y. et al. Characterization of *Mycobacterium leprae* genotypes in China— identification of a new polymorphism C251T in the 16S rRNA gene. *PLoS ONE* 10, e0133268 (2015).

31. Kim, J. P. SNP genotypes of *Mycobacterium leprae* isolated in Korea. *Korean Lepr. Bull.* 45, 3–19 (2012).

32. McLaughlin, R. *Rome and the Distant East: Trade Routes to the Ancient Lands of Arabia, India and China* (A&C Black, 2010).

33. Lavania, M. et al. Single nucleotide polymorphisms typing of *Mycobacterium leprae* reveals focal transmission of leprosy in high endemic regions of India. *Clin. Microbiol. Infect.* 19, 1058–1062 (2013).

34. Lavania, M. et al. Genotyping of *Mycobacterium leprae* strains from a region of high endemic leprosy prevalence in India. *Infect. Genet. Evol.* 36, 256–261 (2015).

35. Das, M., Chaitanya, V. S., Kanmani, K., Rajan, L. & Ebenezer, M. Genomic diversity in *Mycobacterium leprae* isolates from leprosy cases in South India. *Infect. Genet. Evol.* 45, 285–289 (2016).

36. Fontes, A. N. B. et al. Genotyping of *Mycobacterium leprae* present on ZiehlNeelsen-stained microscopic slides and in skin biopsy samples from leprosy patients in different geographic regions of Brazil. *Mem. Inst. Oswaldo Cruz.* 107, 143–149 (2012).

37. Inskip, S. A. et al. Osteological, biomolecular and geochemical examination of an early Anglo-Saxon case of lepromatous leprosy. *PLoS ONE* 10, e0124282 (2015).

38. Mendum, T. A. et al. *Mycobacterium leprae* genomes from a British medieval leprosy hospital: towards understanding an ancient epidemic.

BMC Genom. 15, 270 (2014).

39. Taulman, J. F. & Robbins, L. W. Recent range expansion and distributional limits of the nine-banded armadillo (Dasypus novemcinctus) in the United States. J. Biogeogr. 23, 635–648 (1996).

40. Bouckaert, R. et al. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537 (2014).

41. Phetsuksiri, B. et al. SNP genotypes of Mycobacterium leprae isolates in Thailand and their combination with rpoT and TTC genotyping for analysis of leprosy distribution and transmission. Jpn J. Infect. Dis. 65, 52–56 (2012).

42. Nakata, N., Kai, M. & Makino, M. Mutation analysis of mycobacterial rpoB genes and rifampin resistance using recombinant Mycobacterium smegmatis. Antimicrob. Agents Chemother. 56, 2008–2013 (2012).

43. Bahrmand, A. R., Titov, L. P., Tasbiti, A. H., Yari, S. & Graviss, E. A. High-level rifampin resistance correlates with multiple mutations in the rpoB gene of pulmonary tuberculosis isolates from the Afghanistan border of Iran. J. Clin. Microbiol. 47, 2744–2750 (2009).

44. Comas, I. et al. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat. Genet. 44, 106–110 (2012).

45. DeBarber, A. E., Mdluli, K., Bosman, M., Bekker, L. G. & Barry, C. E. Ethionamide activation and sensitivity in multidrug-resistant Mycobacterium tuberculosis. Proc. Natl Acad. Sci. USA 97, 9677–9682 (2000).

46. Baulard, A. R. et al. Activation of the pro-drug ethionamide is regulated in mycobacteria. J. Biol. Chem. 275, 28326–28331 (2000).

47. Pattyn, S. R. & Colston, M. J. Cross-resistance amongst thiambutosine, thiacetazone, ethionamide and prothionamide with Mycobacterium leprae. Lepr. Rev. 49, 324–326 (1978).

48. Hedge, J. & Wilson, D. J. Practical approaches for detecting selection in microbial genomes. PLoS Comput. Biol. 12, e1004739 (2016).

49. Kai, M. et al. Characteristic mutations found in the ML0411 gene of Mycobacterium leprae isolated in Northeast Asian countries. Infect. Genet. Evol. 19, 200–204 (2013).

50. Zharkov, D. O. Base excision DNA repair. Cell. Mol. Life Sci. 65, 1544–1565 (2008).

51. Jiang, D., Hatahet, Z., Blaisdell, J. O., Melamede, R. J. & Wallace, S. S. Escherichia coli endonuclease VIII: cloning, sequencing, and overexpression of the nei structural gene and characterization of nei and nei nth mutants. J. Bacteriol. 179, 3773–3782 (1997).

52. Saito, Y. et al. Characterization of endonuclease III (nth) and endonuclease VIII (nei) mutants of Escherichia coli K-12. J. Bacteriol. 179, 3783–3785 (1997).

53. Suvarnapunya, A. E. & Stein, M. A. DNA base excision repair potentiates the protective effect of Salmonella Pathogenicity Island 2 within macrophages. Microbiol. Read. Engl. 151, 557–567 (2005).

54. Moolla, N., Goosens, V. J., Kana, B. D. & Gordhan, B. G. The contribution of Nth and Nei DNA glycosylases to mutagenesis in Mycobacterium smegmatis. DNA Repair 13, 32–41 (2014).

55. Ford, C. B. et al. Mycobacterium tuberculosis mutation rate estimates from different lineages predict

substantial differences in the emergence of drugresistant tuberculosis. Nat. Genet. 45, 784–790 (2013).

56. Linder, J. A., Huang, E. S., Steinman, M. A., Gonzales, R. & Stafford, R. S. Fluoroquinolone prescribing in the United States: 1995 to 2002. Am. J. Med. 118, 259–268 (2005).

57. Malik, S., Willby, M., Sikes, D., Tsodikov, O. V. & Posey, J. E. New insights into fluoroquinolone resistance in Mycobacterium tuberculosis: functional genetic analysis of gyrA and gyrB mutations. PLoS ONE 7, e39754 (2012).

58. Yokoyama, K. et al. Impact of amino acid substitutions in B subunit of DNA gyrase in Mycobacterium leprae on fluoroquinolone resistance. PLoS Negl. Trop. Dis. 6, e1838 (2012).

59. Gagneux, S. et al. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science 312, 1944–1946 (2006).

60. Mariam, D. H., Mengistu, Y., Hoffner, S. E. & Andersson, D. I. Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. Antimicrob. Agents Chemother. 48, 1289–1294 (2004).

61. de Vos, M. et al. Putative compensatory mutations in the rpoC gene of rifampin-resistant Mycobacterium tuberculosis are associated with ongoing transmission. Antimicrob. Agents Chemother. 57, 827–832 (2013).

62. Cheng, Y.-S. & Sacchettini, J. C. Structural insights into Mycobacterium tuberculosis Rv2671 protein as a dihydrofolate reductase functional analogue contributing to para-aminosalicylic acid resistance. Biochemistry 55, 1107–1119 (2016).

63. Zheng, J. et al. Para-aminosalicylic acid is a prodrug targeting dihydrofolate reductase in Mycobacterium tuberculosis. J. Biol. Chem. 288, 23447–23456 (2013).

64. Jopling, W. H. & Ridley, D. S. Vadrine (S. 131) in the treatment of lepromatous leprosy: a preliminary report. Lepr. Rev. 29, 143–147 (1958).

65. Honoré, N., Perrani, E., Telenti, A., Grosset, J. & Cole, S. T. A simple and rapid technique for the detection of rifampin resistance in Mycobacterium leprae. Int. J. Lepr. Mycobact. Dis. 61, 600–604 (1993).

66. Singh, P. et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of Mycobacterium lepromatosis. Proc. Natl Acad. Sci. USA 112, 4459–4464 (2015).

67. Singh, P. et al. Genome-wide re-sequencing of multidrug-resistant

68. Mycobacterium leprae Airaku-3. Clin. Microbiol. Infect. 20, O619–O622 (2014).

69. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

70. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359 (2012).

71. Koboldt, D. C. et al. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576 (2012).

72. Rimmer, A. et al. Integrating mapping-, assemblyand haplotype-based approaches for calling variants in clinical sequencing applications. Nat. Genet. 46, 912–918 (2014).

73. Kiełbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487–

493 (2011).

74. Matsuoka, M. The history and characteristics of isolates maintained at the leprosy research center. Nihon Hansen. Gakkai Zasshi 79, 247–256 (2010).

75. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729 (2013).

76. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31, 2745–2747 (2015).

Acknowledgements

We thank the Genomic Technologies Facility at the University of Lausanne for Illumina sequencing and technical support and all the patients and clinical staff who participated in the study. This work was supported by the Fondation Raoul Follereau, the Swiss National Science Foundation grant IZRJZ3_164174, the Swiss Cooperation and Development Center (CODEV), the Heiser Program of the New York Community Trust for Research in Leprosy (grant numbers P15-000827 and P16-000976), and grants CNPq 428964/2016-8 and CAPES PROAMAZONIA 3288/2013.

Author contributions

S.T.C., P.S., C.A., and A.B. designed the study. C.A., P.S., S.G., A.N.B.F., and P.B. processed the samples, extracted DNA, and prepared sequencing libraries. A.B. and C.L. processed the data. A.B. analyzed the data and prepared figures and tables. A.B., C.A., and S.T.C. interpreted the results and wrote the manuscript with input from other authors. S.G., Y.M., M.N., K.B., C.G.S., M.B.S., R.C.B., M.A.C.F., F.B.F., J.G.B., J.A.C.N., S. B.-S., A.L., A.R.A.-S., Y.A.-Q., A.S.A., G.B., L.V.C., F.S., C.R.J., M.Ko., A.F., S.O.S., M.G., O.K., M.M.A.S., G.O.P., P.N.S., E.N.S., M.O.M., P.S.R., I.M.F.D.B., J.S.S., A.A., M.M., and M.Ka. participated in identification of leprosy cases, patient management, sample collection and preparation, and microscopy.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-02576-z.

Supplementary Note

Dating analysis

Dataset

Concatenated SNPs for each sample were used for tip dating analysis. We omitted strains with mutations in the *nth* gene because of their hypermutated genomes, which would interfere with the clock models in the Bayesian inference. Also, we removed highly mutated genes associated with drug resistant strains (**Table 1**) because they might contain mutations that arose from artificial selection during antibiotic treatments. Analysis was done in BEAST2 v2.4.4 [1]. Sites with missing data were included in the analysis because they contain valuable evolutionary signals and missing data are properly handled in BEAST2[2]. Constant (invariable) sites were also included as these contribute to better estimates of population size and branch lengths (https://groups.google.com/forum/#!topic/beast-users/-67kIqZEJf8). We included only unambiguous constant sites, i.e. loci where the reference base was called in all samples. These corresponded to A=230627, C=275179, G=279428, T=231694. Each BEAST run had between 50 M and 100 M MCMC steps to assure convergence and high ESS values for all indicators.

Substitution model

We used the BEAST package bModelTest v0.3.3[3] to infer the best substitution model that fits our data. Strict Clock and Constant Coalescent were used as clock and tree models. Best support resulted for the "123124" model after exploring all reversible models, or only the "transition/transversion split" models (**Supplementary Figure 4**). The "123124" model (A↔C = C↔G and A↔G = C↔T, **Supplementary Figure 5**) was used in subsequent BEAST analyses, although using the simpler HKY or the more complex GTR model produced similar results. bModelTest also supported invariable sites and gamma rate heterogeneity, so "hasGammaRates" and "hasInvariableSites" were used in all BEAST analyses.

Clock model

Mutation rates are not expected to vary significantly among closely related intraspecies taxa, so using the Strict Clock model is generally recommended for such cases[2]. To test whether a strict clock model should be rejected we performed an analysis in BEAST2 under the Log-Normal Relaxed Clock model and two population models, Constant Coalescent and Exponential Coalescent models. The mean coefficient of variation of the relaxed clock was 0.2035 (95% HPD 0.083-0.3256) under the Constant Coalescent model, and 0.1845 (95% HPD 0.0586-0.3086) under the Exponential Coalescent. In both cases, the marginal posterior distribution of the coefficient of variation of the relaxed clock extended down to zero, which means that the strict clock cannot be rejected for our data[2].

Root age under different models

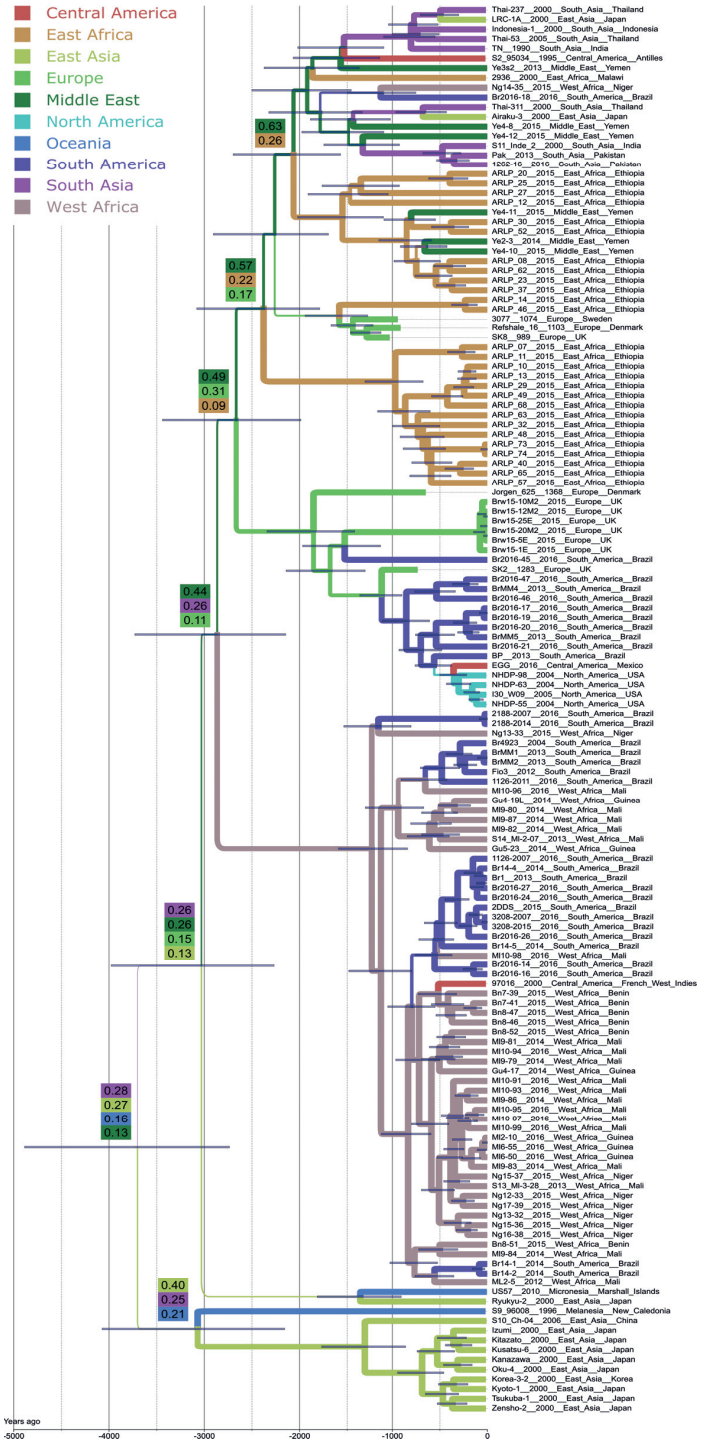
To check for consistency of the dating analysis, we compared the estimates of the divergence time of TMRCA for all *M. leprae* strains under different clock and substitution models. Results remained consistent (**Supplementary Table 1**), indicating that the dataset has a strong clock signal and estimates are not affected by model priors.

***M. leprae* population size through time**

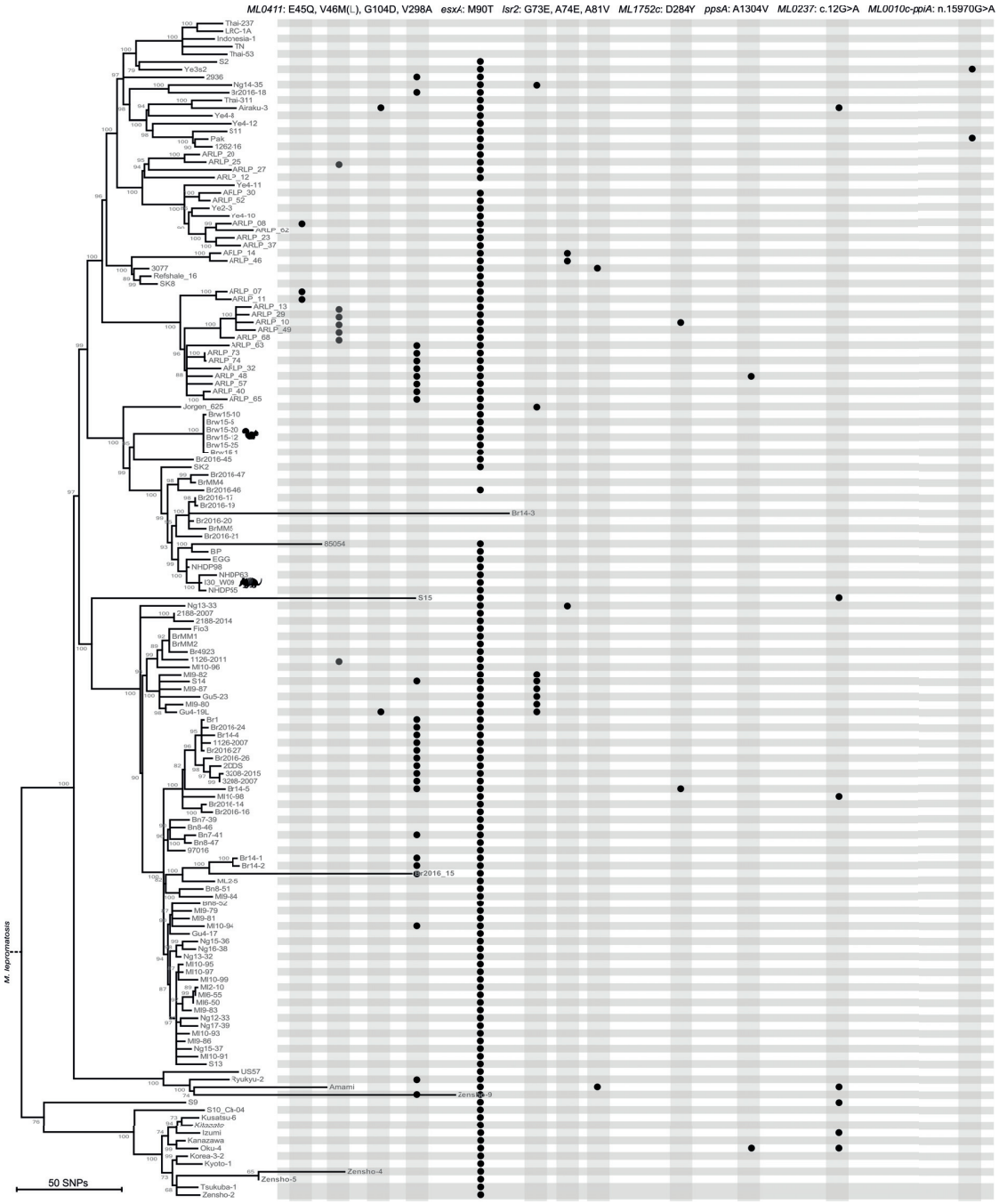
We used the Bayesian Skyline plot (BSP)[4] and the extended Bayesian skyline plot (EBSP)[5] in BEAST2 to infer changes in *M. leprae* population size through time. For the BSP, five and ten dimensions were used, which resulted in overlapping results. BSP resulted in a modest population size change of *M. leprae*, with no evidence of bottlenecks (**Supplementary Figure 6**). The first and last population size intervals in the Bayesian skyline analysis did not overlap, indicating population changes, but the separation was relatively small (4766 – 92708 vs. 94 – 4570) suggesting only a modest change of *M. leprae* population size through time. EBSP did not support a change of population size through time, which is not surprising for single locus data and relatively weak signals of population sizes in the data[2].

Supplementary Table 1 Estimated divergence times for the TMRCA for all *M. leprae* strains under different models in BEAST2.

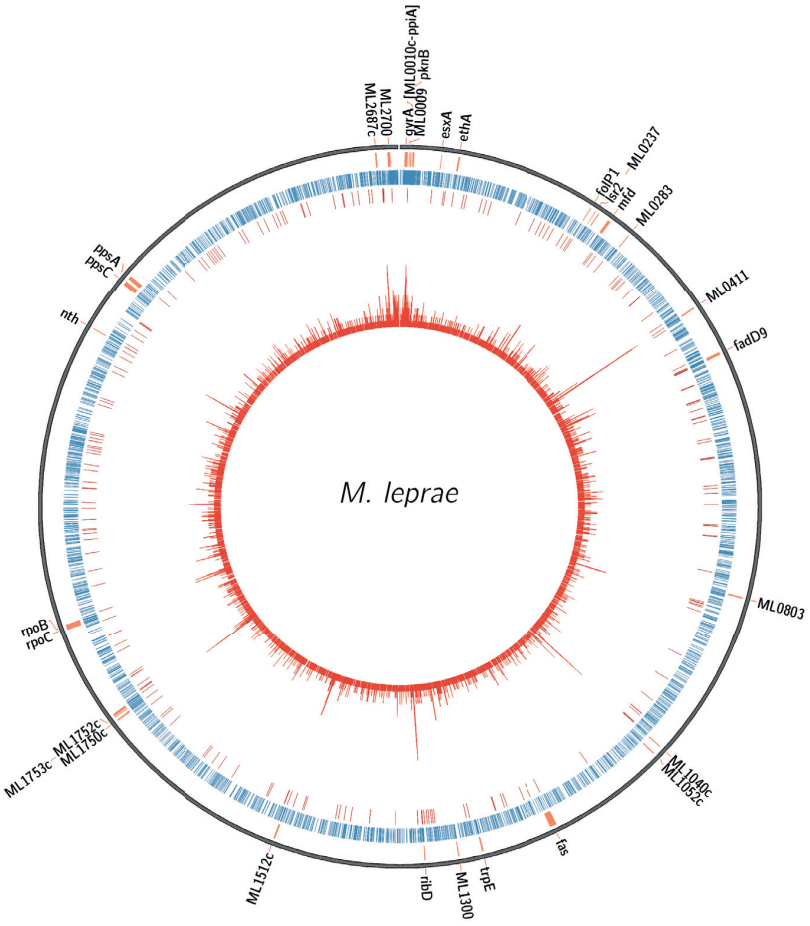
Substitution model	Clock model	Population model	Mean root age (years ago)	95% HPD interval (years ago)
123124	Strict	Coalescent Constant	3697	2693 – 4878
123124	Strict	Coalescent Constant (with Discrete phylogeography)	3699	2731 – 4838
123124	Strict	Coalescent Exponential	3459	2568 – 4467
123124	Strict	Coalescent Bayesian Skyline (5 dimensions)	3544	2694 – 4497
123124	Strict	Coalescent Bayesian Skyline (10 dimensions)	3489	2692 – 4418
123124	Strict	Coalescent Bayesian Skyline Extended	3872	2814 – 5181
123124	Relaxed Log Normal	Coalescent Constant	3774	2492 – 5354
123124	Relaxed Log Normal	Coalescent Exponential	3395	2419 – 4525
bModelTest	Strict	Coalescent Constant	3702	2723 – 4844
HKY	Strict	Coalescent Constant	3692	2648 – 4797
HKY	Strict	Coalescent Exponential	3462	2553 – 4459
GTR	Strict	Coalescent Constant	3684	2685 – 4859
GTR	Strict	Coalescent Exponential	3445	2546 – 4436



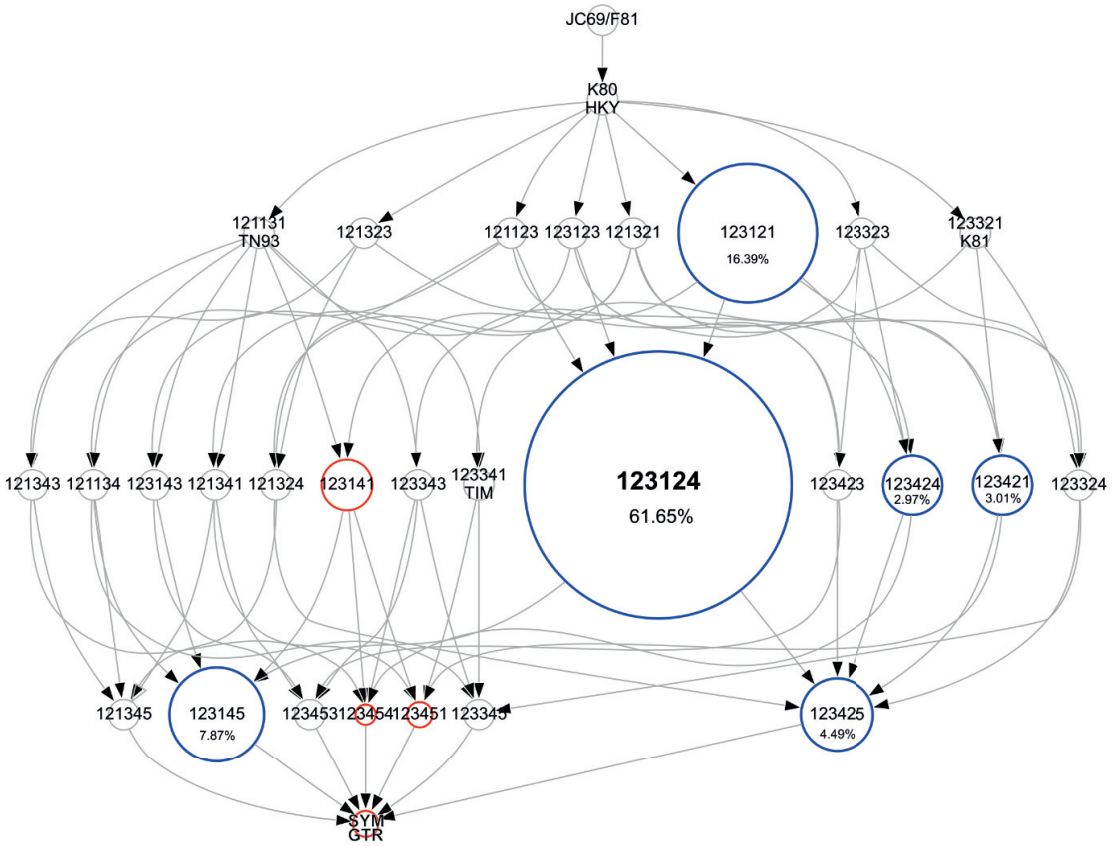
Supplementary Figure 1 Bayesian phylogenetic tree of 146 genomes of *M. leprae* calculated with BEAST 2.4.4. Hypermutated samples with mutations in the *nth* gene were excluded from the analysis. The tree is drawn to scale, with branch lengths representing years of age. Blue horizontal bars show the 95% Highest Posterior Density range of the age for each node. Samples were binned according to geographic origin as given in the legend. Location probabilities of nodes were inferred by the Discrete Phylogeny model and represented by line thickness, and values for the main basal nodes (note that the Discrete Phylogeny model can be influenced by sampling bias[2], therefore results should be interpreted with caution).



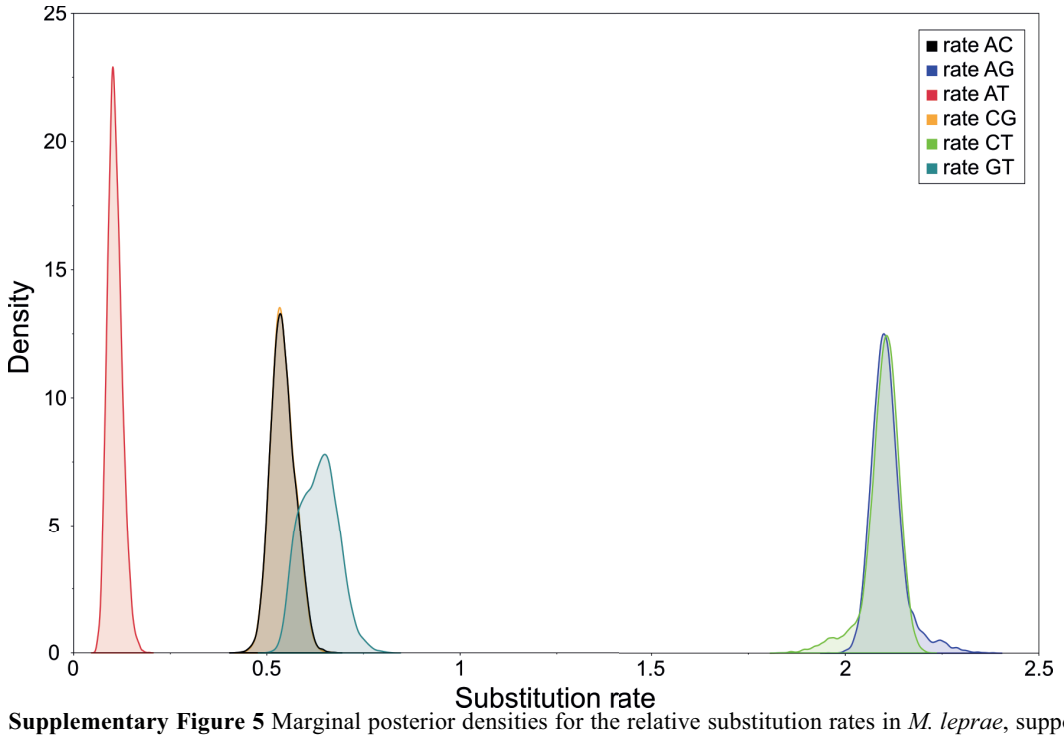
Supplementary Figure 2 Homoplasic sites detected in this study, excluding polymorphic VNTRs and homopolymeric tracts, and genes associated with drug resistance given in **Figure 4** in the main text.



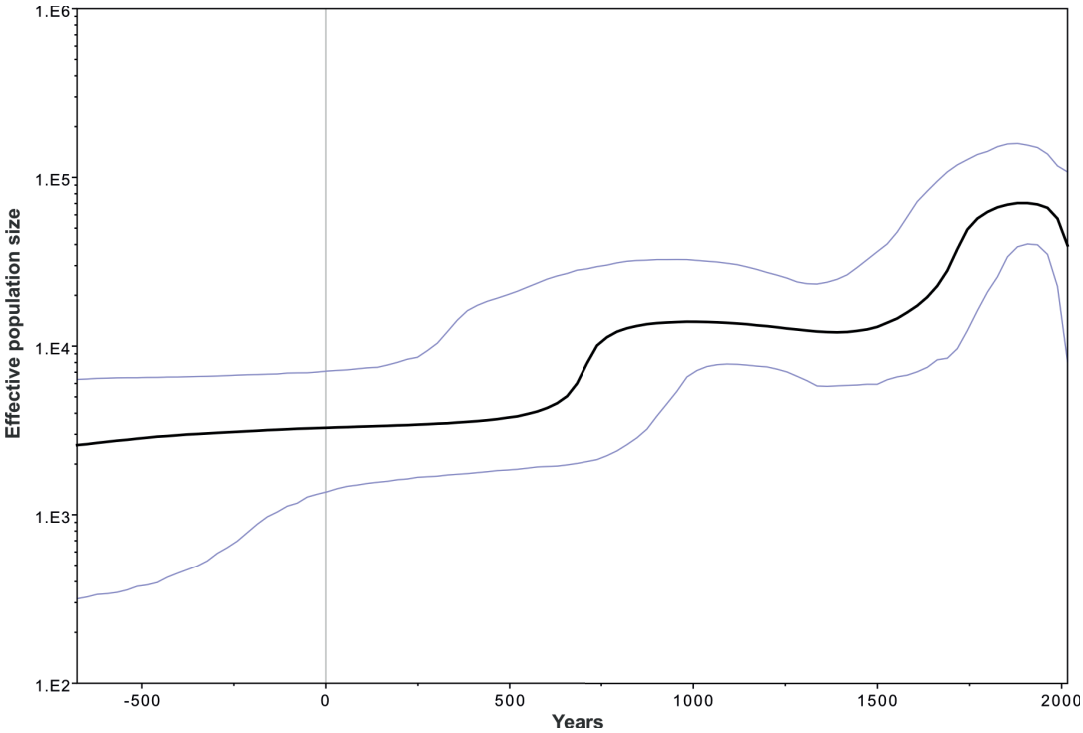
Supplementary Figure 3 Distribution of the mutations found in 154 strains for *M. lepreae*. Outer lane, highly polymorphic genes and genes with homoplasic mutations; 2nd lane (blue), SNPs; 3rd lane (red), InDels; inner lane, variant density per 50-base region.



Supplementary Figure 4 Substitution model evaluation using the bModelTest tool in BEAST2. Models with blue circles are inside 95% HPD, red outside, and grey circles have 0.00% posterior support. Within each circle is the name of the substitution model, while blue circles also show the posterior support.



Supplementary Figure 5 Marginal posterior densities for the relative substitution rates in *M. leprae*, supporting the "123124" substitution model (A↔C = C↔G and A↔G = C↔T). Inferred in BEAST2 using the bModelTest tool.



Supplementary Figure 6 Analysis of the effective population size of *M. leprae* using the Bayesian Skyline plot in BEAST2. Black line is the mean; blue lines indicate the 95% HPD boundaries.

Supplementary References

1. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10,** http://doi.org/10.1371/journal.pcbi.1003537 (2014).

2. Drummond, A. J. & Bouckaert, R. R. *Bayesian evolutionary analysis with BEAST*. (Cambridge University Press, 2015).

3. Bouckaert, R. R. & Drummond, A. J. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evol. Biol.* **17,** 42 (2017).

4. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22,** 1185–1192 (2005).

5. Heled, J. & Drummond, A. J. Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8,** 289 (2008).

Chapter 5. : Animal reservoir

Red squirrels from the British Isles are infected with leprosy bacilli

Charlotte Avanzi,[1]* Jorge del-Pozo,[2]* Andrej Benjak,[1]* Karen Stevenson,[3] Victor R. Simpson,[4] Philippe Busso,[1] Joyce McLuckie,[3] Chloé Loiseau,[1]† Colin Lawton,[5] Janne Schoening,[6] Darren J. Shaw,[2] Jérémie Piton,[1] Lucio Vera-Cabrera,[7] Jesùs S. Velarde-Felix,[7] Fergal McDermott,[6] Stephen V. Gordon,[6,8,9,10] Stewart T. Cole,[1]#Anna L. Meredith[2]#

[1]Global Health Institute, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland.
[2]Royal (Dick) School of Veterinary Studies and Roslin Institute, University of Edinburgh, Easter Bush Campus, Roslin, Scotland, UK.
[3]Moredun Research Institute, Pentlands Science Park, Bush Loan, Edinburgh, Scotland, UK.
[4]Wildlife Veterinary Investigation Centre, Chacewater, Cornwall, UK.
[5]School of Natural Sciences, Ryan Institute, National University of Ireland, Galway, Ireland.
[6]UCD School of Veterinary Medicine, University College Dublin, Belfield, Dublin, Ireland.
[7]Laboratorio Interdisciplinario de Investigación Dermatológica, Servicio de Dermatología, Hospital Universitario, Monterrey, N.L., Mexico.
[8]UCD School of Medicine, University College Dublin, Belfield, Dublin, Ireland.
[9]UCD School of Biomolecular and Biomedical Science, University College Dublin, Belfield, Dublin, Ireland.
[10]UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Belfield, Dublin, Ireland.

* These authors contributed equally to this work
# Corresponding author

2016, Science, 354(6313); 744-747

Contributions: study design, genomic part: sample processing, DNA extraction, library preparation, array capture, interpretation of results, and manuscript preparation

This is a pre-print, author-modified version of an article accepted for publication in Science following peer review. The version of record is available online in free open access at: **doi.org/10.1126/science.aah3783**

Leprosy, caused by infection with *Mycobacterium leprae* or the recently discovered *Mycobacterium lepromatosis*, was once endemic in humans in the British Isles. Red squirrels in Great Britain (*Sciurus vulgaris*) have increasingly been observed with leprosy-like lesions on the head and limbs. Using genomics, histopathology, and serology, we found *M. lepromatosis* in squirrels from England, Ireland, and Scotland, and *M. leprae* in squirrels from Brownsea Island, England. Infection was detected in overtly diseased and seemingly healthy animals. Phylogenetic comparisons of British and Irish *M. lepromatosis* with two Mexican strains from humans show that they diverged from a common ancestor around 27,000 years ago, whereas the *M. leprae* strain is closest to one that circulated in Medieval England. Red squirrels are thus a reservoir for leprosy in the British Isles.

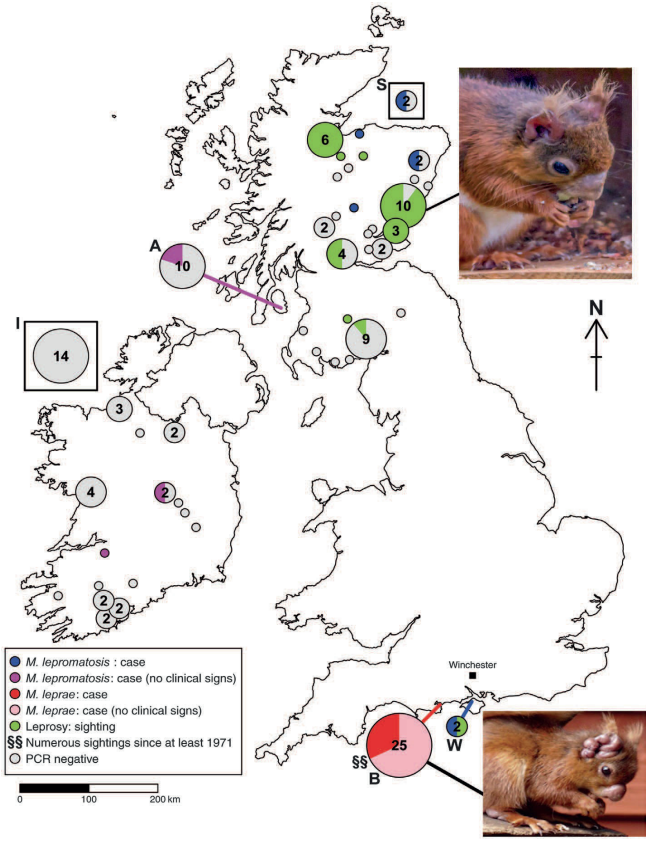


Figure 1: Squirrel sampling sites in the British Isles. Pie charts indicate the location of sites where squirrels were sighted or found and are color-coded as indicated in the box; numbers within circles indicate different animals tested where $N > 1$. Boxed circles refer to squirrels of unknown location. I, Ireland; S, Scotland; A, Isle of Arran; B, Brownsea Island; W, Isle of Wight. The figure was drawn in R (v3.2.23) with the package *maps* (v3.1.0) using the *mapdata* (v2.2-6) "worldHiresMapEnv" and the package *plotrix* (v3.6-2) for pie charts.

Often considered a disease of the past, leprosy remains a public health problem in certain lowand middle-income countries, with ~220,000 new cases reported annually (*1*). Leprosy was rife in Europe in the 15th and 16th centuries, probably because of social segregation, other infectious diseases such as plague, or changes in host immunity (*2–5*). Today, all British clinical cases occur in individuals with a history of residence in a leprosy-endemic country (*6*). The disease manifests indifferent forms, ranging from multibacillary (or lepromatous) to paucibacillary (or tuberculoid), depending on the immunogenetics of the host (*4*). In all forms, skin lesions are accompanied by peripheral nerve damage, which causes sensory loss and may lead to deformities.

It had generally been accepted that leprosy resulted solely from interhuman transmission of *Mycobacterium leprae*. But in recent years, compelling evidence emerged from the south-ern United States for zoonotic cases after exposure to infected nine-banded armadillos (*Dasypus novemcinctus*) (*7–9*). Furthermore, *M. leprae* was considered to be the sole causative agent of leprosy until 2008, when a new species, *Mycobacterium lepromatosis*, was identified in patients with diffuse lepromatous leprosy (DLL) (*10*). Such cases were primarily associated with Mexico and the Caribbean region (*11*). Comparison of the genome sequences of *M. lepromatosis* and *M. leprae* revealed that

despite separating millions of years ago, the two genomes are remarkably similar in their size, organization, and (pseudo)gene content, but show only 88% sequence identity (*11*).

The Eurasian red squirrel *Sciurus vulgaris* is a widespread Palearctic species found from Ireland in the west to Kamchatka in the east (*12*, *13*). However, in the United Kingdom, the*S. vulgaris* population of ~140,000 is severely threatened by habitat loss, squirrel poxvirus infection, and competition with >2.5 million gray squirrels, *Sciurus carolinensis*, introduced from North America (*14*, *15*). Because of their endangered status, red squirrels are now protected (*16*). Recent detection of mycobacterial infection in red squirrels was reported in Scotland, with lesions and histopathology characteristic of DLL and evidence for *M. lepromatosis* being the etiological agent (*17*). Similarly affected squirrels were observed on the Isle of Wight and Brownsea Island in southern England (*18*), and observations of squirrel leprosy in Scotland are increasing (Fig. 1). Here, we investigated the leprosy outbreak using 70 red squirrel cadavers from Great Britain, with or without disease signs; 40 cadavers from Ireland, where no sightings of squirrels with leprosy signs have been reported; and four Scottish gray squirrel cadavers.

A differential polymerase chain reaction (PCR) screen was implemented to detect *M. leprae* and *M. lepromatosis* DNA (*11*). We analyzed a total of 172 tissue samples from 13 animals with leprosy features and 101 without leprosy features (tables S1 and S2) (*19*). Six Scottish squirrels (two without clinical signs) (*17*), two from Ireland (no clinical signs), and one from the Isle of Wight, England (*18*) contained *M. lepromatosis* in several tissue samples from different anatomical sites, whereas all 25 red squirrels (17 without clinical signs) tested from

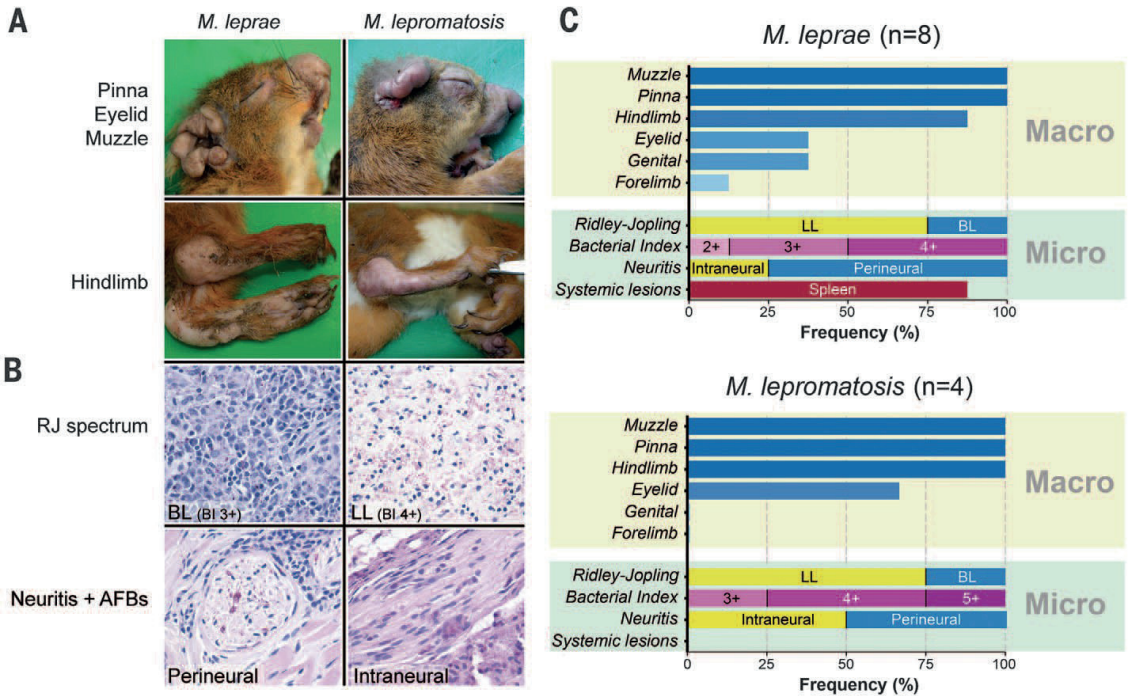


Figure 2: Gross histopathological features of red squirrels with leprosy. (A) Macroscopic features of squirrels infected with *M. lepromatosis* or *M. leprae* are similar. (B) Histological examination of tissue sections from infected squirrels using the Ridley-Jopling (RJ) classification after Ziehl Neelsen staining (magnification 400×). LL, lepromatous leprosy; BL, borderline lepromatous leprosy. (C) Summary of main macroscopic and microscopic findings from squirrels infected with *M. leprae* (n = 8) or M. *lepromatosis* (n = 4).

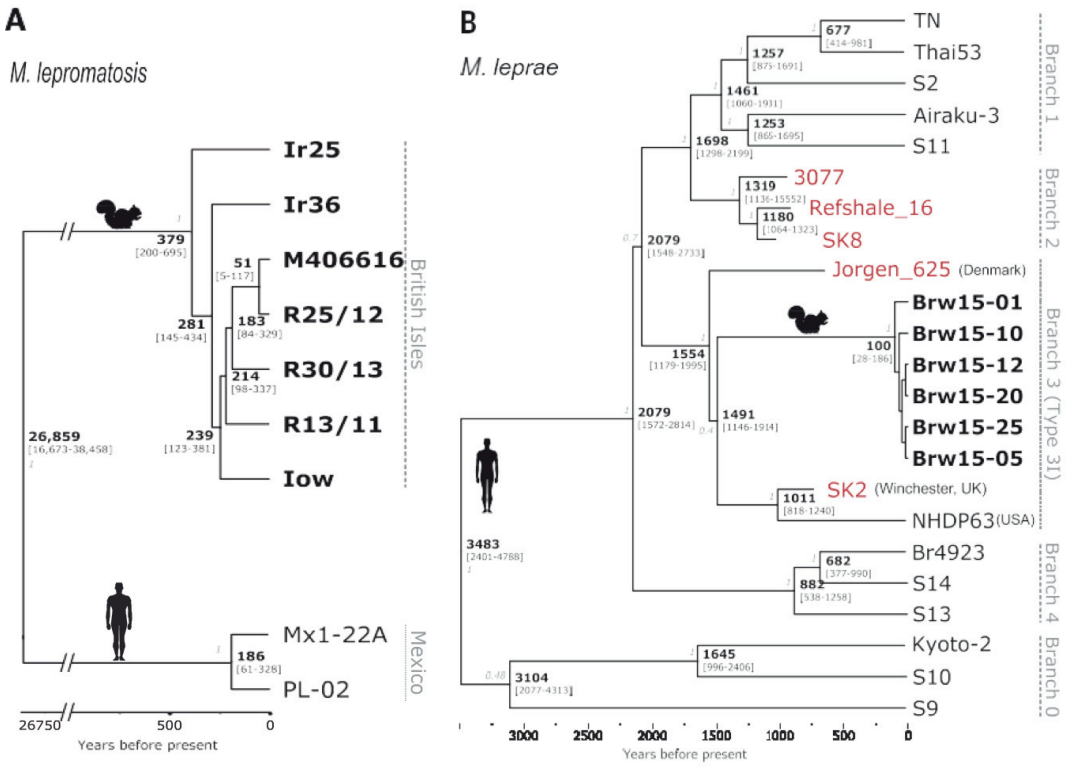


Figure 3: Phylogeny of leprosy bacilli. (A) Bayesian phylogenetic tree representation of nine *M. lepromatosis* genome sequences obtained from squirrels (bold) or humans (upper and lower parts, respectively), calculated by BEAST 1.8.2 (*24*) using the mutation rate of *M. leprae* and inferred from 432 genome-wide variable positions. Squirrel sample prefixes: Ir, Ireland; Iow, Isle of Wight; all others from Scotland. Both human strains were from Mexico. (B) Bayesian phylogenetic tree representation of *M. leprae* inferred from 498 genome-wide variable positions, calculated as in (A). For squirrel samples (bold), Brw denotes Brownsea Island cluster; red labeling indicates ancient strains for which radiocarbon dating information was available (*3*). For both trees, divergence time intervals are shown on each node in years before present, with the 95% highest posterior density (HPD) range in brackets. Posterior probabilities for each node are shown in gray.

Brownsea Island were infected with *M. leprae* (Fig. 1 and table S3). No cases of co-infection were observed (table S3). From the combined results, we concluded that 21% [21/101; 95% confidence interval (CI), 13 to 30%] of the squirrels without clinical signs and all 13 of the animals with clinical signs harbored leprosy bacilli.

Serological tests were performed on nine diseased and 14 healthy red squirrels from Scotland and England, as well as the four gray squirrels. The grays were all seronegative, whereas 13 of 23 blood samples from red squirrels contained antibodies for the leprosy-specific antigen PGL-1 (phenolic glycolipid–1) (*20*) (table S4). Serology is useful to confirm the disease and to predict infection in live animals but cannot be used for species identification, as both *M. leprae* and *M. lepromatosis* produce this cell wall antigen (*11*). Diseased Scottish squirrels infected with *M. lepromatosis* displayed a range of macroscopic lesions, including alopecia and extensive swelling of the snout, lips, eyelids, ear pinnae, and limb extremities (Fig. 1, Fig. 2A, fig. S1, and tables S2 and S5) (*19*). Histopathological examination of four such squirrels (Fig. 2B) revealed granulomatous dermatitis, sheets of epithelioid macrophages, and large numbers of acid-fast bacilli (AFB). There was neural involvement with the presence of AFB in nerve endings; neuritis was patchy and more frequently perineural (Fig. 2C). Inflammation was not focused exclusively around nerves and was mostly dermal. There were no signs of vasculitis, but AFB were present intravascularly (Fig. 2C). Similar lesions were observed in eight squirrels from Brownsea Island infected with *M. leprae*, although these animals also harbored numerous AFB in the spleen (Fig. 2C). Overall, the macroscopic signs and histopathology were characteristic of

lepromatous leprosy (Fig. 2, A and B, and figs. S2 and S3). From post mortem inspection of diseased squirrels, it was not possible to distinguish between infection with *M. lepromatosis* or *M. leprae*, as is the case in human leprosy (*11*, *21*, *22*).

To obtain deeper insight into the strains responsible and to perform phylogenetic analyses, we used a variety of DNA enrichment techniques (table S6) prior to Illumina sequencing, which was necessary because neither *M. leprae* nor*M. lepromatosis* can be cultured (*19*). Sufficient sequence coverage of *M. lepromatosis* genomes from seven squirrels was obtained (table S7). In parallel, we sequenced an additional genome of*M. lepromatosis*, Pl-02, froma PGL-1–seropositive patient from Sinaloa, Mexico (tables S1 and S4). The resultant sequence reads were mapped against the reference *M. lepromatosis* genome sequence from a patient from Monterrey, Mexico (*11*) to identify polymorphisms. Consistent with previous *M. leprae* genome comparisons (*9*, *11*, *23*), there was an exceptionally high level of sequence conservation between *M. lepromatosis* strains (99.99% identity) despite their different geographic origins. The two Mexican patient isolates differed by only seven single-nucleotide polymorphisms (SNPs), whereas the number of SNPs in the six British and Irish strains ranged from 1 to 17 on pairwise comparisons (table S8). Overall, there are roughly 400 SNPs that distinguish *M. lepromatosis* strains from Mexico and the British Isles (table S8). Clustering of Mexican and British *M. lepromatosis* strains into two distinct lineages was supported by maximumparsimony (fig. S4) and neighbor-joining (fig. S5) phylogenetic reconstructions. On the basis of the *M. leprae* mutation rate (*19*) and using the Bayesian inference software BEAST (*24*), we estimated that the British Isles and Mexican strains diverged from their most recent common ancestor around 27,000 years ago, whereas the Irish and British strains diverged as recently as 200 years ago (Fig. 3A). The latter estimate is consistent with the date of the first campaign to reintroduce the red squirrel into Ireland from England between 1820 and 1856, following its extinction in the 17th century (*12*, *25*). This suggests that these animals may already have been infected with *M. lepromatosis* when they were reintroduced. Finding *M. leprae* in red squirrels in England was unexpected, because leprosy was eradicated from the British Isles several centuries ago, thus demonstrating that a pathogen can persist in the environment long after its clearance from the human reservoir. Furthermore, this is only the second report of *M. leprae* in nonprimate species. From Bayesian and maximum-parsimony analysis (Fig. 3B and fig. S4A), we note that the two closest relatives to the strain of *M. leprae* found on Brownsea Island were both from medieval Europe. Intriguingly, one of these (SK2) originated from the skeletal remains of a leprosy victim buried about 730 years ago in Winchester, a city situated a mere 70 km from Brownsea Island (Fig. 1). Like SK2, the Brownsea Island strain of *M. leprae* belongs to sequence type 3I, which forms a distinct *M. leprae* branch (Fig. 3B) (*3*) and is now endemic in wild armadillos in the southern United States (*9*). Thus, *M. leprae* with this particular sequence type is capable of infecting at least three different hosts: humans, red squirrels, and armadillos. Because there were no obvious genomic polymorphisms restricted to the *M. leprae* 3I type that might account for this broad host range (tables S9 and S10), we explored the possibility that these three species might share a major susceptibility gene and focused on *TLR1*. This candidate gene, encoding the surface-exposed Toll-like receptor 1 (TLR1) displayed on various

epithelial and immune cells, is known to be associated with susceptibility to leprosy (Fig. 4A). A dysfunctional *TLR1* allele encoding an Ile^{602} → Ser variant with an altered transmembrane domain is prevalent in Caucasians and is asso ciated with a decreased risk for leprosy (*5*, *26*). By contrast, the *TLR1* Asn^{248} → Ser variant is associated with an increased risk of leprosy in humans. This mutation is located in the ninth repeat of the extracellular leucine-rich repeat (LRR) region of TLR1 (*27*). Furthermore, in ninebanded armadillos, an Arg^{627} → Gly change in TLR1 [close to the Toll-interleukin receptor (TIR) domain; Fig. 4A] seemingly confers resistance to leprosy (*28*). Using PCR, the coding exon of *TLR1* was amplified and sequenced from 58 red squirrels (with or without lesions) and three gray squirrels (tables S11, S12, and S14) (*19*). Upon comparison of the sequences and TLR1 alignments (table S13), no polymorphisms were observed at the same sites associated with leprosy in humans and armadillos. However, in some red squirrels, two distinct polymorphic sites exist: a single SNP leading to a Ser^{494} → Asn mutation in the 19th repeat of the LRR region, and a cluster of linked mutations that produce Ser^{657} → Asn (S657N), Leu^{660} → Val (L660V), and Asn^{662} → Cys (N662C) variants in helix 1 of the TIR domain (Fig. 4B). These mutations were found less frequently in squirrels infected with leprosy bacilli than in healthy animals, which suggests that they may confer protection (for S494N, odds ratio = 5.77; 95% CI, 1.42 to 23.41; P = 0.01; for S657N-L660V-N662C, odds ratio = 4.89; 95% CI, 0.98 to 24.53; P = 0.05).

It is unclear whether leprosy is contributing to the demise of the red squirrel population or how these animals became infected with *M. lepromatosis* or *M. leprae*. Because *M. lepromatosis* has

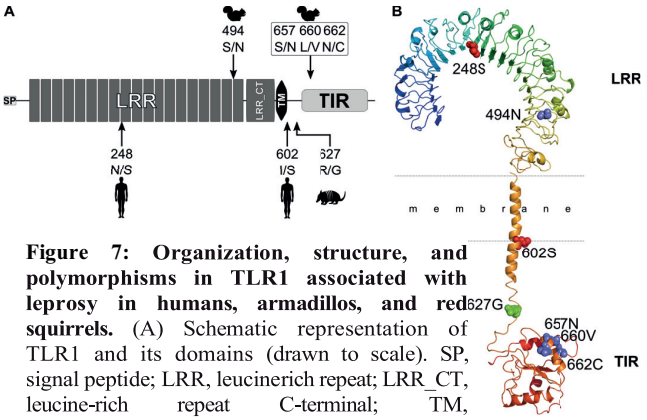


Figure 7: Organization, structure, and polymorphisms in TLR1 associated with leprosy in humans, armadillos, and red squirrels. (A) Schematic representation of TLR1 and its domains (drawn to scale). SP, signal peptide; LRR, leucinerich repeat; LRR_CT, leucine-rich repeat C-terminal; TM, transmembrane domain; TIR, Toll/interleukin1 receptor.

(B) Structural model of the red squirrel TLR1. Protein is colored in a rainbow spectrum from N terminus (blue) to C terminus (red). Amino acid abbreviations: C, Cys; G, Gly; I, Ile; L, Leu; N, Asn; R, Arg; S, Ser; V, Val.

only recently been discovered as a human pathogen (*10*) and few detailed case reports have been published (*10*, *11*, *21*, *29*), further investigation is required to establish its relative prevalence in wildlife compared to humans. *M. leprae* was long considered to be an obligate human pathogen that was introduced to the Americas by European settlers, prior to anthroponotic infection of armadillos, because there are no human skeletal remains with signs of leprosy from the pre-Columbian era (*9*). The discovery that the strain of *M. leprae* in red squirrels on Brownsea Island today is essentially the same as one that circulated in medieval England and Denmark, and is highly related to the extant North American armadillo strain, raises the possibility of a second anthroponotic introduction in Europe. If this were the case, it must have occurred several centuries ago, as leprosy became increasingly scarce in the British Isles after the 17th century (*3*). It is also conceivable that humans may have been infected through contact with red squirrels bearing *M. leprae*, as these animals were prized for their fur and meat in former times (*30*). Our findings show that further surveys of animal reservoirs of leprosy bacilli are warranted, because zoonotic infection from such reservoirs may

contribute to the inexplicably stubborn plateau in the incidence of the human leprosy epidemic despite effective and widespread treatment with multidrug therapy (*1*).

REFERENCES

1. World Health Organization, *Wkly. Epidemiol. Rec.* 88, 365–379 (2013).
2. H. D. Donoghue *et al*., *Infect. Genet. Evol.* 31, 250–256 (2015).
3. V. J. Schuenemann et al., *Science* 341, 179–183 (2013).
4. A. Alter, A. Grant, L. Abel, A. Alcaïs, E. Schurr, Mamm. *Genome* 22, 19–31 (2011).
5. S. H. Wong et al., *PLOS Pathog*. 6, e1000979 (2010).
6. N. Fulton, L. F. Anderson, J. M. Watson, I. Abubakar, *BMJ Open 6*, e010608 (2016).
7. R. Sharma et al., *Emerg. Infect. Dis*. 21, 2127–2134 (2015).
8. R. Truman, *Lepr. Rev*. 76, 198–208 (2005).
9. R. W. Truman et al., *N. Engl. J. Med*. 364, 1626–1633 (2011).
10. X. Y. Han et al., *Am. J. Clin. Pathol*. 130, 856–864 (2008).
11. P. Singh et al., *Proc. Natl. Acad. Sci*. U.S.A. 112, 4459–4464 (2015).
12. M. Carey, G. Hamilton, A. Poole, C. Lawton, The Irish Squirrel Survey 2007 (COFORD, Dublin, 2007).
13. S Harris, G. B. Corbet, The Handbook of British Mammals (Mammal Society/Blackwell Scientific, ed. 3, 1991).
14. D. M. Tompkins, A. W. Sainsbury, P. Nettleton, D. Buxton, J. Gurnell, *Proc. R. Soc. B* 269, 529–533 (2002).
15. E. Stokstad, Science 352, 1268–1271 (2016).
16. Council of Europe, Convention on the Conservation of European Wildlife and Natural Habitats (ETS No. 104), Appendix III (1979). https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680304356
17. A. Meredith et al., *Vet. Rec*. 175, 285–286 (2014).
18. V. Simpson et al., *Vet. Rec*. 177, 206–207 (2015).
19. See supplementary materials on Science Online.
20. J. S. Spencer, P. J. Brennan, *Lepr. Rev*. 82, 344–357 (2011).
21. J. S. Velarde-Félix, G. Alvarado-Villa, L. Vera-Cabrera, *Am. J. Trop. Med. Hyg*. 94, 483–484 (2016).
22. L. Vera-Cabrera et al., J. *Clin. Microbiol*. 53, 1945–1946 (2015).
23. M. Monot et al., *Nat. Genet*. 41, 1282–1289 (2009).
24. A. J. Drummond, A. Rambaut, *BMC Evol. Biol*. 7, 214 (2007). 38, 49–58 (2015).
25. S. R. Krutzik et al., *Nat. Med*. 9, 525–532 (2003).
26. C. de Sales Marques et al., *J. Infect. Dis*. 208, 120–129 (2013).
27. L. B. Adams et al., Mem. Inst. Oswaldo Cruz 107 (suppl. 1), 197–208 (2012).
28. P. G. Jessamine et al., J. *Drugs Dermatol*. 11, 229–233 (2012).
29. P. Lurz, Red Squirrel: Naturally Scottish (Scottish Natural Heritage, 2010).

ACKNOWLEDGMENTS

We thank E. Sheehy, E. Goldstein, M. Flaherty, A. Zintl, the National Trust, Forestry Commission Scotland, and Saving Scotland's Red Squirrels for samples, help, and advice. We thank the Genomic Technologies Facility at the University of Lausanne for Illumina sequencing and technical support. Raw sequence read files have been

SUPPLEMENTARY MATERIALS

Available online at www.sciencemag.org/content/354/6313/744/suppl/DC1

Chapter 6. : Diagnostics

Evaluation of auramine O Staining and conventional PCR for leprosy diagnosis: a comparative cross-sectional study at Armauer Hansen Research Institute, Addis Ababa, Ethiopia

Selfu Girma[1#], Charlotte Avanzi[2], Kidist Bobosha[1], Kassu Desta[2], Munir Idris[1], Philippe Busso[2], Yohannes Tsegaye[1], Shimelis Nigusse[3], Tsegaye Hailu[1], Stewart T. Cole[2#a], Abraham Aseffa[1]

[1]Armauer Hansen Research Institute
[2] Global Health Institute, École Polytechnique Fédérale de Lausanne, Switzerland
[3]Addis Ababa University, CHS, Department of Medical Laboratory Sciences
[4] ALERT hospital
[#a] Institut Pasteur de Paris, France

Corresponding author

2018, manuscript in preparation

Contributions: genomic part: sample processing, DNA extraction, PCR, laboratory training, manuscript preparation

Abstract

Background

Diagnosis of leprosy mainly relies on clinical examination due to the inconsistent sensitivity and poor reproducibility of the current laboratory tests. Utilisation of alternative methods to the standard Ziehl Neelsen (ZN), Fite-Faraco (FF) and Haematoxylin and Eosin (H&E) staining procedures may eventually improve leprosy diagnosis.

Methodology/ Principal findings

In this comparative study, the performance of the fluorescent Auramine O (AO) staining and polymerase chain reaction (PCR) was assessed with different skin samples using a combination of ZN, FF and H&E staining as the gold standard. AO, ZN, FF, H&E and PCR tests were performed on slit skin smears (SSS) and/or punch biopsies collected from 141 clinically confirmed leprosy cases and 28 non-leprosy skin samples. DNA was extracted from punch biopsies using two different methods with or without mechanical lysis.

Sensitivities were 87.6%, 59.3% and 77% for H&E, ZN and FF, respectively, whereas they reached 65.5% and 77.9% for AO in SSS and tissue sections and 91.1% for PCR in tissue sections. Also, among samples with low bacillary index, sensitivities of AO staining (58.6%) was similar to FF (56.9%, $p > 0.05$) and lower than PCR (79.3%, $p < 0.05$). Sensitivity of PCR also increased (96.8%, $p < 0.05$) when mechanical lysis was used during DNA extraction compared to enzymatic treatment alone (86.7%).

Conclusions/Significance

Our results showed that for diagnostic purposes, analysis of skin section is more sensitive than SSS, especially for samples with a low bacillary load. AO staining on SSS and tissue sections was not significantly better than other routine diagnostic tests but was considerably more user friendly. The sensitivity of PCR was higher than current standard methods and increased when combined with more efficient DNA extraction using mechanical and chemical lysis. Therefore, we recommend AO staining for the diagnosis of leprosy in lower health institutions and PCR diagnosis at the referral level.

Authors' Summary

Leprosy is one of the oldest reported infectious diseases in the history of humanity, with a remarkably high rate of active transmission still reported annually, but mainly in resource-limited endemic countries. The primary tool for leprosy transmission prevention is diagnosing the disease at an early stage. However, this is a challenging activity in resource-limited disease-endemic countries with the currently available microscopic routine diagnostic tests. We found that Auramine O staining can alternatively be used for leprosy diagnosis using light-emitting diode fluorescence microscopy on both slit skin smear and tissue section samples with a potential of replacing the routine light microscopic examination. Also, we showed that a combination of highly efficient DNA extraction with a single run of conventional PCR outperformed standard microscopy methods for all samples with high and low bacillary index.

Introduction

Mycobacterium leprae is the causative agent of leprosy, a chronic granulomatous infectious disease affecting the skin and peripheral nerves (1). Leprosy manifests in various forms based on the immunological profiles and bacterial load in patients (1). According to Ridley and Jopling, leprosy is classified as indeterminate (IND), tuberculoid (TT), borderline tuberculoid (BT), borderline (BB), borderline lepromatous (BL) and lepromatous leprosy (LL) (2). More recently, for therapy purposes, the World Health Organization (WHO) implemented another classification depending on the number of lesions. Patients with <5 skin lesions are considered as paucibacillary (PB) cases and are treated for six months with two antibiotics

whereas those with 6 or more lesions are regarded as multibacillary (MB) and receive three drugs for one year (3, 4).

In 2016, 214,783 new cases of leprosy were reported worldwide (5) including 19,384 (9%) in Africa. With 3,692 new cases, Ethiopia was the African country with the second highest prevalence. The trend of new cases reported for the last ten years is stable with 4,086 per year on average (5). This data indicated the ongoing active transmission despite intense efforts to eliminate leprosy as a public health problem and the widespread use of multidrug therapy (MDT) (6). Lack of reliable diagnostic tools especially for the early stage of the disease is of major concern (7, 8). Hence efforts to improve diagnosis are being undertaken and WHO has also set early detection of leprosy as a priority in leprosy control strategy (9, 10).

Since *M. leprae* cannot be cultivated *in vitro*, clinical signs such as presence of lesions, sensory loss, and thickened peripheral nerves, serve as the primary tool of leprosy diagnosis. However, the disease can easily be confused with other skin pathologies indicating the need for a differential leprosy diagnosis especially by less experienced physicians (3, 5, 11) For this reason, along with clinical examination, identification of the bacteria and the histopathological classification in skin samples is necessary to confirm leprosy diagnosis. The most popular tools are Ziehl-Neelsen (ZN) and Fite-Faraco (FF) staining performed on clinical samples such as slit-skin smears (SSS), nasal swabs and formalin fixed paraffin embedded (FFPE) tissue samples (12-16). Even though ZN and FF are available at lower level health institutions of resource-limited countries, their performance in detecting *M. leprae* bacilli is low, particularly in PB patients (17). Therefore, for these problematic cases, the clinician mostly relies on clinical examination which requires experience. In return, this highlights the need for more sensitive techniques to support clinical diagnosis. Auramine O (AO) staining is a fluorescence-based method widely used to detect mycobacterial species such as *M. tuberculosis* and *M. leprae* (18, 19). AO has been previously evaluated to be more sensitive for *M. leprae* detection in tissue sections compared to FF and is less time-consuming (20, 21). Molecular methods such as conventional PCR are even more sensitive, and can help with leprosy diagnosis (13, 22-25). However, such techniques are not widely available. Hence, this study was designed to determine the diagnostic utility of AO staining and conventional PCR in routine diagnosis in comparison with the standard protocol.

Methods

Ethical Consideration

Ethical clearance was obtained from the AHRI/ALERT Ethical Review Committee, Addis Ababa University College of Health Science, Department of Medical Laboratory Science Ethics and Research Review Committee and Ministry of Science and Technology, National Research Ethics Review Committee. Written informed consent was obtained from participants and parents or guardians of participating children.

Study population

A total of 141 leprosy cases comprising 136 newly diagnosed treatment naïve and five relapse leprosy patients with any form of the disease, were enrolled in this comparative cross-sectional study at the All African Leprosy, Tuberculosis and Rehabilitation (ALERT) center from January 2015 to April 2016. All cases were clinically diagnosed and confirmed by a dermatologist. Non-leprosy patients (n=28) visiting the minor surgery department of the ALERT hospital were enrolled in the study as a control group.

Data and sample collection

Sociodemographic and clinical parameters

Nurses collected sociodemographic data and clinical information of study participants using a structured questionnaire at the ALERT Red Medical Clinic before the participants went to

the sample collection area. Participants in the control group were recruited before their admission for routine surgical treatment.

Sample collection

Slit skin smears (SSS) from were collected from three different body sites (right and left earlobes, and either an eyebrow, the forehead or one of the arms) of leprosy patients to increase the probability of detecting acid-fast bacilli. While collecting SSS for routine ZN diagnosis, a duplicate slide was prepared from the same site at the same time and sent to the AHRI pathology laboratory for AO staining (26).

One 6 mm skin punch sample was collected by well-trained nurses for each leprosy patient. Punch biopsy collection was not performed on cosmetic and sensitive body parts like the face and scrotal area (27, 28). For the non-leprosy control group, skin biopsy samples of ~10 mm were collected from discarded skin specimens after routine surgical treatment. After collection, each punch biopsy was divided into two parts; one was placed in 10% buffered formalin solution to be processed for staining and the other in 70% ethanol for DNA extraction.

Sample processing

ZN and AO staining on SSS

One of the slides containing SSS was dried for 15 min at room temperature and fixed from below by passing slowly through the flame of a spirit burner three times. The slide was then stained with 1% carbolfuchsin solution, heated as above until vapor begins to rise for 5 min. The slide was then washed with running tap water and destained with 1% acid-alcohol for 10-20 s, rinsed with tap water gently, counterstained with 0.2% methylene blue for 1 min, washed again with tap water and air dried. Finally, it was examined under a 100X objective of a conventional light microscope (26).

For AO staining, the slide was flooded with 0.1% AO (MERCK, Germany, prepared locally) solution for 20 min, destained with 0.5 % acid-alcohol for 2 min, counterstained with locally prepared 0.5% potassium permanganate (Riedel-deHaen, Germany) for 4 min, then air dried. The slide was rinsed with sterile water between each step. Then, bacilli examination was carried out using a light-emitting diode fluorescence microscopy (ZEISS Primo Star iLED fluorescence microscope, Germany) with a 40X objective (17, 29).

Tissue processing, embedding, and sectioning

Punch biopsies in 10% formalin were kept for 48-72 h before tissue processing was performed overnight using an automated tissue processor (LEICA ASP 300S, Germany) as explained elsewhere (15). The following day, the tissue was embedded in real-paraffin wax. A series of 4 µm thick tissue sections were prepared using a rotary microtome (LEICA RM2255, Germany) and fixed on one end of a frosted slide coated with 50% egg albumin – glycerol prepared locally. A total of three slides, each containing four consecutive sections from the same tissue were prepared for AO, FF and H&E staining (15).

H&E staining

One of the slides containing tissue sections was deparaffinized and rehydrated using a dry oven at 60^0 C for 30 min and in two changes of xylene for 10 min, in a series of decreasing concentrations of alcohol and finally in tap water. The slide was stained for 8 min in Harris's hematoxylin reagent, destained in 0.5% acid-alcohol for 3 s and washed in running tap water. It was then counterstained with 0.5% eosin for 1 min. After dehydration with alcohol and xylene treatment, it was mounted with DPX mounting medium for histopathologic examination to be examined by a pathologist. Histopathologic features of leprosy like granulomas, epithelioid cells, foamy macrophages, giant cells, type of cell infiltration, and inflammation were used to

diagnose and classify the disease into its different forms (15, 30).

FF and AO staining of tissue sections

Slides were warmed in an oven at 60^0 C for 10 min and deparaffinized twice with two parts xylene and one part of vegetable oil, for 15 min, then blotted well with absorbent paper to remove the xylene-oil remnant and hydrated in a jar containing distilled water. For FF staining, the slide was flooded with filtered 1% calbolfuchsin for 20 min followed by destaining with 10% H_2SO_4 for 2 min. Tissue sections were then counterstained with 0.25% methylene blue solution for 20 s. The slide was rinsed with sterile water between each step. AO staining was performed as outlined above. Finally, for both staining procedures, after a final wash with water, slides were blotted, cleared with xylene, mounted with mounting medium (DPX mountant for histology, Sigma) and examined under the 100X objective of the microscope or using LED-FM under the 40X objective, respectively (28, 31, 32).

DNA extraction

DNA was extracted using three different methods. The first method called host depletion (HD) (33) (Benjak et al, 2017, Nat Comm, accepted), removes host DNA and is therefore mainly used for whole-genome sequencing applications, but it can be also used for PCR application. HD was applied to 35 skin biopsies (S1 Table). The second method used the QIAmp UCP Pathogen Mini kit (QiagenGmbH, Hilden, Germany) with an adapted protocol on 40 skin biopsies (S1 Table). Briefly, biopsies were cut into small pieces in a 1.8 mL micro-centrifuge tube. AHL Lysis buffer (500 µl) containing 20 µl of proteinase K (20 mg/mL) was added to the disrupted tissue and incubated for 1 h at 56°C. After mechanical lysis twice with 200 µl of 0.1 mm zirconium beads (Bertin Technology) at a velocity of 6.5 m/s for 45 s with 5 min incubation on ice in a Precellys® 24 Instrument), a second round of enzymatic lysis was performed using 40 µl of proteinase K (20 mg/mL) prior to incubation with APL2 buffer for 10 min at 70°C. DNA was precipitated and purified on QIAamp UCP Pathogen Mini silica column followed by elution in 100 µl of elution buffer. In the third method, DNA was extracted from 95 skin biopsies (S1 Table) using the QIAmp Fast DNA Tissue kit (QigenGmbH, Hilden, Germany) without mechanical lysis. Briefly, biopsies were cut into small pieces in a 1.8 mL micro-centrifuge tube. AHL Lysis buffer (500 µl) containing 20 µl of proteinase K (20 mg/mL) was added to the disrupted tissue and incubated 1h at 56°C, then for 10 min at 70°C, as above. DNA was then precipitated, purified on QIAamp silica column and eluted in 100 µl of elution buffer.

Polymerase chain reaction (PCR)

Conventional Polymerase Chain Reaction (PCR) was performed using primer pairs to detect the *M. leprae* specific repetitive region RLEP and the specific region in *hemN* from *M. lepromatosis* in *M. leprae* PCR negative samples (33, 34). For each reaction, 3 5 µl of extracted DNA was mixed with each primer (200 nM final), 25 µL of Accustart Master Mix and water in a final volume of 50 µL. Amplification cycles started with a denaturation step at 95°C for 5 min, followed by 40 cycles at 95°C for 30 s, annealing at 58°C for 40 s and extension at 72°C for 30 s. The reaction ended with an additional 10 min extension step at 72°C. The amplified PCR product was then examined by agarose gel (1% w/v) electrophoresis.

Quality Control

Samples known to be positive or negative for *M. leprae* were used as positive and negative controls during the staining procedures and PCR.

Data Analysis

Based on previous experience of similar studies to develop a reference standard (35), we have established a combination of ZN, FF and H&E staining tests for this specific study. Clinical diagnosis was the necessary part of

this combination supported by at least one or more positive test results of H&E, ZN and FF staining. This test panel was chosen due to their routine application in diagnosing leprosy in our laboratory and worldwide.

However, since the specificity of all these methods is known to be low, we will consider samples obtained from non-leprosy patients as truly negative for the specificity of AO staining in tissue, FF and PCR methods. A "true positive" will be a sample with one or more positive test results of H&E, ZN and FF staining. For ZN and AO in SSS, "true negatives" will be the negative samples obtained with the alternative gold standard method since SSS samples were not obtained from the non-leprosy patients (S1 and S2 Tables).

Socio-demographic data, clinical information and laboratory results were introduced into Stata SE version 11 for statistical analysis. Data obtained from four samples of leprosy cases were excluded from analysis due to incompleteness. Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated including 95% confidence intervals (CI) against the designed alternative gold standard (Fig S1). For statistical significance between the different detection methods, a binomial test (MacNemar test or the exact binomial test) and Fisher's test were calculated in R when applied on the same group of samples and in case of independent groups, respectively.

Moreover, in other studies, patients commonly classified as TT, BT and INT are usually considered as PB patients with low BI whereas LL, BL and BB are classified as MB cases with high BI (36). However, the WHO classification is based on the number of skin lesions and is not linked to the R&J classification because BI is either high or low. Moreover, it is also commonly accepted that TT, BT and BB samples are associated with low bacillary index whereas LL, BL and BB have higher bacillary index even if some exceptions can be observed. Thus, the MB and PB classification of this study is only relative to the number of skin lesions found per patient. Nevertheless, to compare the diagnostic performance of the methods described here with others published elsewhere, we have classified the patients based on the R&J classification as follows: TT, BT, IND and negative (NEG) will be considered as low BI samples (LB) and LL, BL and BB as high BI samples (HB).

Results

Socio-demographic characteristics and clinical features

A total of 169 participants were involved in the study from January 2015 to April 2016 at ALERT center, Addis Ababa, Ethiopia. There were 141 leprosy cases and 28 in the non-leprosy control group (S1 and S2 Tables). Male study participants comprised 63.9 % (108/169) with a male to female ratio of 1.7:1. The mean age and SD of study participants was 35. 8 ± 14.6 years with age ranging between 15 and 75 years.

Clinical features

Among the clinically confirmed leprosy cases, 19.9% (28/141) showed ⩽5 skin lesions and about 80.1 % (113/141) presented with >5 skin lesions and were classified accordingly as PB and MB (S1 Table). Visible physical disability was seen in 59.6% of the leprosy patients. Five (3.5%) participants who had completed MDT were categorized as relapse cases based on clinical criteria. A total of 32 (22.5%) participants presented with leprosy reactions classified as pure neuritis 15.6% (5/32), reversal reactions 68.8% (22/32) and 15.6 % (5/32) with erythema nodusum leprosum. Regarding family history, 25.5 % (36/141) of the leprosy cases used to live with a leprosy patient.

Among the non-leprosy control group, 32.1 % (9/28) came to the hospital for surgical treatment of skin cancer, while the remaining 67.9% (19/28) came for different surgical

treatment including corrective amputation but with no history of leprosy (S2 Table).

The gold standard method

Among the 141 clinically confirmed leprosy cases, four samples were excluded from the analysis because of the absence of data for the microscopy methods (S1 Table). A total of 137 clinically confirmed cases were analyzed and 113 were positive according to the gold standard method including 99 positive and 14 negative by histopathology (H&E) (Table 1). A total of 58 and 55 cases belonged to the HB and LB sample groups (S3 Table). All 28 samples from non-leprosy patients were negative using the gold standard method classification.

Performance of leprosy diagnosis using Auramine O staining and PCR

Auramine O staining

On analysis of the 137 SSS, the sensitivity of AO in SSS (65.5%) was slightly higher (p>0.05) than ZN (59.3%) while specificity was 100% for both tests (Table 2, S4 Table).

The sensitivity and specificity of 137 tissue sections stained with FF staining were 77% and 100%, respectively (Table 2), while other statistical parameters, PPV and NPV were 100% and 51.8%, respectively. Sensitivity and specificity of AO-tissue staining are similar (p>0.05) to FF with 77.9% and 100%, respectively, using the established gold standard method (Table 2, S4 Table).

The overall sensivity of both AO in tissue and FF is significantly higher (p<0.05) than AO in SSS and ZN (S4 Table). In addition sensitivity of the different tests is higher in HB compared to LB groups (p<0.05) as expected (Table 3, S4 table).

DNA extraction and PCR

DNA samples extracted from all of the non-leprosy control groups were negative. PCR was positive for 104/113 gold standard positive leprosy cases. All DNA samples that were PCR-negative for *M. leprae* were also PCR-negative *M. lepromatosis*.

The global sensitivity and specificity of the method were 91.1% and 100%, respectively (Table 2). Also, the performance of different DNA extraction methods (with or without mechanical lysis) was compared. For samples where DNA was extracted with mechanical lysis (HD and Qiamp UCP Pathogen kit), the overall sensitivity is statistically higher (96.8%, p<0.05) compared to samples where DNA was extracted without mechanical lysis (Qiamp fast Pathogen) with 86.7% (Table 2, S4 Table). The disparity between the two methods mainly occurs in the LB group (BT, TT, IND and NEG) with a sensitivity of 95.5% with mechanical lysis and 76.7% without mechanical lysis (p>0.05) (Table 3, S4 Table).

Diagnostic performance of laboratory test when leprosy cases are histopathologically classified and confirmed

Among the 58 (51.3%) histopathologically confirmed samples collectively classified under BB, BL or LL and expected to have high bacterial concentration, all laboratory tests gave similar results ranging from 84.5% to 98.3% with the highest sensitivity (p<0.05) recorded for PCR (Table 2, S4 Table). On the other hand, of the 55 (48.7%) samples classified as BT, TT, INT or Neg and thus expected to have few or no bacilli count (S1 Table), the sensitivity of AO in SSS (43.1%) is slightly higher than ZN (31%, p>0.05) whereas the sensitivity of AO in tissue (58.6%) is similar to FF (56.9%, p>0.05) but statistically lower than PCR (79.3%, p<0.05) (Fig. 1, Table 3 and Fig S1).

Negative cases

A total of 24 cases with clinical signs of leprosy were considered negative using the gold standard method. Regarding the number of lesions, 11 and 13 patients were classified as MB and PB, 17 presented with disabilities and nine reported a family history of leprosy (S1 Table). While ZN, FF and AO in SSS showed negative results, AO in tissue and PCR were

positive for six and ten cases, respectively, including four positive samples common to both methods (Fig. 1, S3 Table).

Discussion

Current leprosy diagnosis relies upon clinical examination of the patient, recognition of skin lesions and peripheral neuropathy, in addition to identification of acid-fast bacilli and histopathology typical of the active lesion. However, the identification of a true leprosy case when disabilities are not yet visible, especially for PB patients, is a challenge for the clinician (25). Therefore, histopathology is still mostly used as the gold standard for the diagnosis, with some limitations (25). This method is less specific compared to the mycobacterial staining methods, but different reports describe a valuable sensitivity of histopathologic analysis for some doubtful cases (37). In this study, 72% (99/137) of the clinically identified leprosy patients would have been considered positive based solely on H&E staining and clinical signs (Table 1). Using a combination of methods, as suggested by Reja *et al,* including in this case the H&E staining and FF on tissue with ZN on SSS, increased the number of positives to 82% (113/137) (25).

ZN is uncomplicated, cost-effective and the most frequently used method for the detection of AFB especially in resource limited settings. The sensitivity of ZN is inconsistent ranging from 18% to 56%, depending on the study (23-25, 38). We reported a sensitivity of 59.3% respectively, with a low negative predictive value demonstrating the probable high rate of false negative for ZN. An acceptable alternative would be AO staining on SSS with a slightly higher percentage of positivity (64.9%, $p>0.05$). While the difference is not significant, AO staining is simpler due to the ease of detection of fluorescently stained bacilli and the ability to screen the entire field within a short period (Fig. 2).

FF staining is another widely-accepted laboratory diagnostic test for leprosy on tissue sections. Though its specificity is usually high as suggested by our results and others (20), the sensitivity of FF is affected by the type of disease, as are most of the other laboratory tests for leprosy. Nayak et al. reported an FF sensitivity of 44.6% and 60%, respectively for PB and MB patients, whereas we report 61.8% and 93.1% for LB (BT, TT, IND and NEG) and HB (LL, BL and BB) cases, respectively (20). The ALERT hospital is specialized in diagnosis of dermatological diseases and has many senior dermatologists. In this study, the difference between the sensitivity value is most probably linked to the definition of the gold standard method and the involvement of highly skilled dermatologists. Here, FF and AO is tissue are more sensitive than ZN ($p<0.05$) and the global sensitivity between FF (77%) and AO (77.9%) in tissue are similar. The detection rate obtained for AO staining is similar to that in previously published studies (29) Moreover, sensitivity is identical between both methods for HB and LB cases suggesting that AO staining on the tissue (Fig. 2) could replace FF without any loss of sensitivity. In addition, the sensitivity of both FF and AO in tissue section is higher ($p<0.05$) compared to SSS for LB cases. This suggests that tissue sections should be preferred to SSS for leprosy diagnosis.

PCR is often acknowledged for its great sensitivity among all laboratory diagnostic tests (22, 23). A study in Brazil reported PCR sensitivity of 40 % for TT, 55.5 % for BT and 100 % for all BB, BL and LL cases, respectively (39). The authors concluded that PCR improves the diagnostic efficiency of LB cases which mostly have a negative BI (39). In our study, the sensitivity was 98.3% for LL, BL and BB (HB) cases and 83.6% for BT, TT, IND and NEG (LB) cases. The result for HB samples was comparable with that of the Brazilian study but the sensitivity found for the LB samples was relatively higher. Even though we were not able to confirm independently the histopathologic

classification, we emphasize that the possible reason for this higher sensitivity is the use of a more effective DNA extraction method. *M. leprae* is an intracellular pathogen with an elaborate cell wall which confers resistance to alcohol and acid treatment as well as to standard pathogen lysis methods. Altogether, these characteristics should be taken into consideration to ensure proper DNA recovery. The importance of the extraction method used to obtain *M. leprae* DNA is often underestimated. Indeed, in this study, we detected more positive cases when chemical lysis was combined with mechanical lysis during DNA extraction with an increase of sensitivity ($p>0.05$) from 86.7% to 96.8% compared to DNA extraction using chemical lysis alone (Table 3).

Finally, 10/24 PCR samples among the negative cases, classified by the gold standard method established here, were positive by PCR for which specificity was 100% in our investigation. In previous studies, false positives have been observed in samples from patients with other skin diseases but this was probably due to misdiagnosis in the first place (39). To avoid false positives, only patients with no family history of leprosy were included in the non leprosy control group and all skin samples were analyzed with standard methods such as H&E and FF. Thus, the rate of positivity in the negative gold standard group is highly encouraging to recommend PCR even for routine diagnosis. Overall, these results indicate the potential value of a single run of PCR to support clinical diagnosis rapidly without the requirement of pathologists and the other staining tests included in the alternatively establish the gold standard method in the study. However, drawback of conventional endpoint PCR is non-quantitative nature. Currently, several quantitative PCR tests have been optimized for detection of *M. leprae* but the cost and the absence of a standardized protocol is a limitation to its implementation at lower level health institutions in resource-limited countries (39-41).

Acknowledgements

We would like to acknowledge Ms. Cecile Prebandier for her kind support. We would also like to mention the great support from ALERT Red Medical Clinic, minor surgery unit, ALERT laboratory and dermatology department during recruitment of participants and sample collection. Our special thanks will also go Sr. Genet Amare and Sr. Haregewion Yetesha their quality management approach and for handling the sample collection.

Figures and tables

Table 1: Number and repartition of positive samples according to the gold standard method

Histopathology (H&E)	Total	FF positive – ZN negative	FF negative – ZN positive	FF and ZN positive	FF and ZN negative
Positive	99	19	2	57	21
Negative	38	6	3	5	24

Table 2: Diagnosis performance of the laboratory tests with 95% confidence interval (CI) based on the establish gold standard method – $PCR_{m+e:}$ PCR result obtained on DNA extracted with mechanical (m) and enzymatic (e) methods; PCR_e: : PCR result obtained on DNA extracted with enzymatic (e) method only

Starting material		Sensitivity		PPV		NPV	
		%	CI (95%)	%	CI (95%)	%	CI (95%)
SSS	ZN	59.3	49.668.4	100	94.6-100	34.3	23.346.6
	AO	65.5	56-74.2	100	95.1-100	38.1	26.1-51.2
Punch biopsy	H&E	87.6	80.193.1	100	96.3-100	66.7	50.580.4
	FF	77	68.184.4	100	95.8-100	51.8	37.865.7
	AO	77.9	69.1-85.1	100	95.9 - 100	52.8	38.6-66.7
	PCR	91.1	84.395.7	100	96.5-100	73.7	56.9 77.4
	PCR_{m+e}	96.8	88.7 – 99.6	100	93.9100	95	83.1 – 99.4
	PCR_e	84.6	71.993.1	100	92100	80	64.690.9

Table 3: Sensitivity of tests for different histopathological classified groups with high BI (HB) and low BI (LB) – $PCR_{m+e:}$ PCR result obtained on DNA extracted with mechanical (m) and enzymatic (e) methods; PCR_e: : PCR result obtained on DNA extracted with enzymatic (e) method only

Starting material		HB (LL, BL, BB) (n=47)		LB (BT, TT and NEG) (n=55)	
		Sensitivity	CI (95%)	Sensitivity	CI (95%)
SSS	ZN	84.5	78.7-90.2	31.0	25.3-36.8
	AO	87.9	82.2-93.7	43.1	37.5-48.9
Punch biopsy	FF	93.1	87.3-98.8	56.9	51.18-62.6
	AO	93.1	87.3-98.8	58.6	52.9-64.4
	PCR	98.3	92.5-100	79.3	73.5-85.1
	PCR_{m+e}	100.0	92.7-100	92.0	83.2-100
	PCR_e	95.5	86.1-100	76.7	68.7-84.7

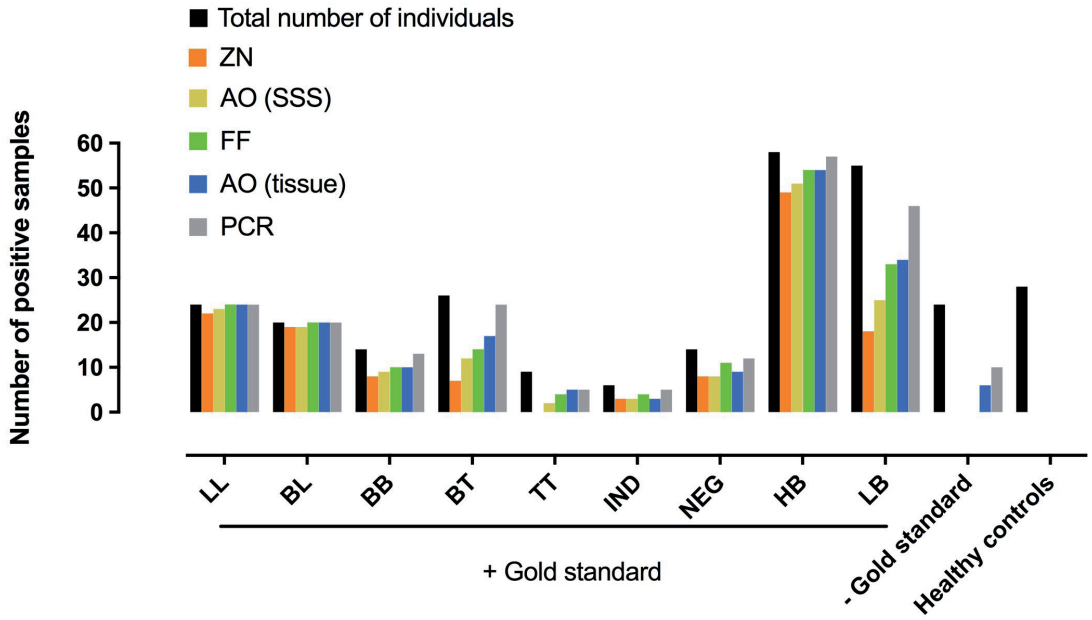


Figure 8: Histopathological repartition of the gold standard positive and negative samples with the number of positive sample and positivity rate (%) for each laboratory tests – HB: high bacillary load included samples from the LL, BL and BB groups; LB: low bacillary load included samples from the BT, TT, IND and NEG groups – The graphic show the high positivity rate for AO-tissue and PCR in all groups compared to other methods.

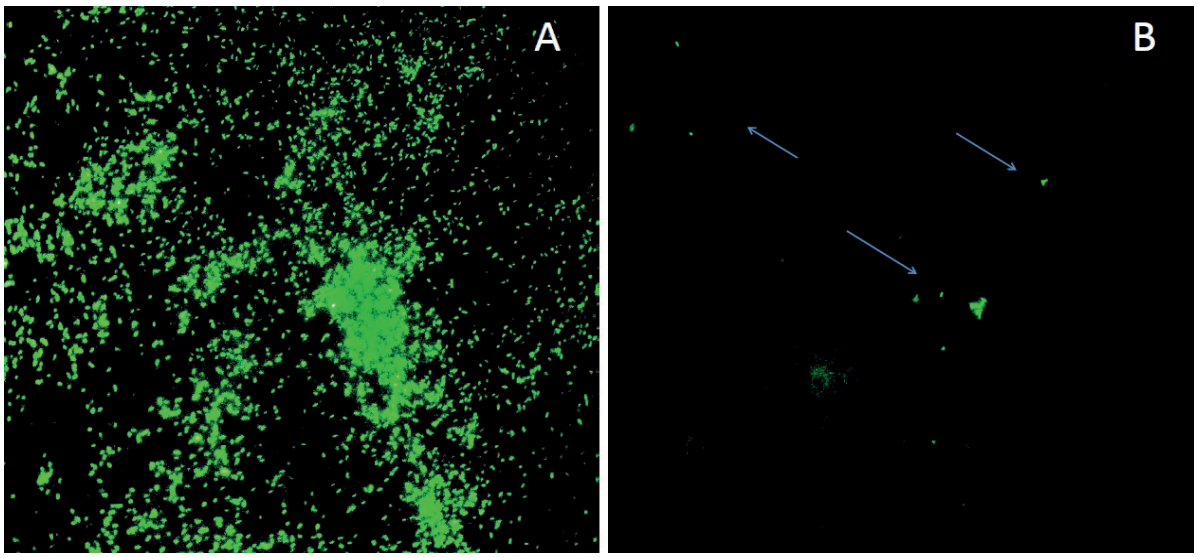


Figure 2. Auramine O stained *M.leprae* in FFPE tissue section under 40X objective of light-emitting diode fluorescence microscope A: Sample with high bacillary load (HB) B. Sample with low bacillary load (LB).

Reference

1. Britton WJ, Lockwood DN. Leprosy. Lancet (London, England). 2004;363(9416):1209-19.

2. Ridley DS, Jopling WH. A classification of leprosy for research purposes. Leprosy review. 1962;33:119-28.

3. Chaka B, G/Yesus D, Tsegaye F, Mosisa F, Lemecha G, Kebede T, et al. Tuberculosis, Leprosy and TB/HIV Prevention and Control Programme. 4 ed: Federal Ministry of Health Ethiopia; 2008.

4. WHO. Drug use in Leprosy : Model Presribing Information 1988.

5. WHO. Weekly epidemiological record. Geneva: 2017.

6. WHO. Global Leprosy Strategy 2016–2020 -Accelerating towards a leprosy-free world. 2016.

7. Goulart IM, Goulart LR. Leprosy: diagnostic and control challenges for a worldwide disease. Archives of dermatological research. 2008;300(6):269-90.

8. Souza VNBd. Challenges for laboratory diagnosis of leprosy Hansen Int. 2011;36(2):7-8.

9. WHO. Global Strategy for Further Reducing the Leprosy Burden and Sustaining Leprosy Control Activities. 2005.

10. Geluk A, Bobosha K, van der Ploeg-van Schip JJ, Spencer JS, Banu S, Martins MV, et al. New biomarkers with relevance to leprosy diagnosis applicable in areas hyperendemic for leprosy. Journal of immunology (Baltimore, Md : 1950). 2012;188(10):4782-91.

11. Moschella SL, Garcia-Albea V. Diffrential Diagnosis of Leprosy. International text book of leprosy 2016.

12. Kumaran SM, Bhat IP, Madhukara J, Rout P, Elizabeth J. Comparison of bacillary index on slit skin smear with bacillary index of granuloma in leprosy and its relevance to present therapeutic regimens. Indian journal of dermatology. 2015;60(1):51-4.

13. Lastoria JC, Abreu MA. Leprosy: a review of laboratory and therapeutic aspects-part 2. Anais brasileiros de dermatologia. 2014;89(3):389-401.

14. Singh PA, Agarwal R, Misra V, Gupta SC, Bajaj AK. Clinico-histopathological concordance in leprosy. Tropical doctor. 2000;30(4):228-31.

15. Banerjee S, Biswas N, Kanti Das N, Sil A, Ghosh P, Hasanoor Raja AH, et al. Diagnosing leprosy: revisiting the role of the slit-skin smear with critical analysis of the applicability of polymerase chain reaction in diagnosis. International journal of dermatology. 2011;50(12):1522-7.

16. Patrocinio LG, Goulart IM, Goulart LR, Patrocinio JA, Ferreira FR, Fleury RN. Detection of *Mycobacterium leprae* in nasal mucosa biopsies by the polymerase chain reaction. FEMS immunology and medical microbiology. 2005;44(3):311-6.

17. Adiga DSA, Hippargi SB, Rao G, Saha D, Yelikar BR, Karigoudar M. Evaluation of Fluorescent Staining for Diagnosis of Leprosy and its Impact on Grading of the Disease: Comparison with Conventional Staining. Journal of

Clinical and Diagnostic Research. 2016;10(10).

18. Ryan GJ, Shapiro HM, Lenaerts AJ. Improving acid-fast fluorescent staining for the detection of mycobacteria using a new nucleic acid staining approach. Tuberculosis (Edinburgh, Scotland). 2014;94(5):511-8.

19. Hooja S, Pal N, Malhotra B, Goyal S, Kumar V, Vyas L. Comparison of Ziehl Neelsen & Auramine O staining methods on direct and concentrated smears in clinical specimens. The Indian journal of tuberculosis. 2011;58(2):72-6.

20. Nayak SV, Shivarudrappa AS, Mukkamil AS. Role of fluorescent microscopy in detecting *Mycobacterium leprae* in tissue sections. Annals of diagnostic pathology. 2003;7(2):78-81.

21. N. A, Nagarajappa A, Prabhu D. Sensitivity Of Fluorescent Microscopy In Detecting *Mycobacterium leprae* In Tissue Sections. The Internet Journal of Pathology. 2010;11(2).

22. Scollard DM, Adams LB, Gillis TP, Krahenbuhl JL, Truman RW, Williams DL. The Continuing Challenges of Leprosy. Clinical Microbiology Reviews. 2006;19(2):338-81.

23. Soto A, Muñoz PT. Leprosy Diagnosis: An Update on the Use of Molecular Tools Lucrecia. Molecular Biology. 2015;4(4).

24. Siwakoti S, Rai K, Bhattarai NR, Agarwal S, Khanal B. Evaluation of Polymerase Chain Reaction (PCR) with Slit Skin Smear Examination (SSS) to Confirm Clinical Diagnosis of Leprosy in Eastern Nepal. PLoS neglected tropical diseases. 2016;10(12).

25. Reja AH, Biswas N, Biswas S, Dasgupta S, Chowdhury IH, Banerjee S, et al. Fite-Faraco staining in combination with multiplex polymerase chain reaction: a new approach to leprosy diagnosis. Indian journal of dermatology, venereology and leprology. 2013;79(5):693-700.

26. Groenen G, Saunderson P, Ji B. ILEP Learning Guide three : How to do a skin smear examination for leprosy. London: International Federation of Anti-Leprosy Associations; 2003. p. 3.

27. Levitt J, Bernardo S, Whang T. How to Perform a Punch Biopsy of the Skin. N ENG J MED. 2013;369(11).

28. Nischal U, Nischal KC, Khopkar U. Techniques of Skin Biopsy and Practical Considerations. Journal of Cutaneous and Aesthetic Surgery. 2008 Jul-Dec;1(2):107-11. PubMed PMID: PMC2840913.

29. Bhardwaj K, Ghate S, Dhurat R. Detection of *Mycobacterium leprae* in tissue sections using auramine O fluorescent stain versus modified fite-faraco: A comparative study. International Journal of Infectious Diseases. 2016;45:388.

30. Pires CA, Miranda MF, Bittencourt Mde J, Brito AC, Xavier MB. Comparison between histopathologic features of leprosy in reaction lesions in HIV coinfected and non-coinfected patients. Anais brasileiros de dermatologia. 2015;90(1):27-34.

31. WHO. Fluorescent light-emitting diode (LED) microscopy for diagnosis of tuberculosis : Policy Statment. 2011.

32. Chaidir L, Parwati I, Annisa J, Muhsinin S, Meilana I, Alisjahbana B, et al. Implementation of LED

fluorescence microscopy for diagnosis of pulmonary and HIV-associated tuberculosis in a hospital setting in Indonesia. PloS one. 2013;8(4):61727.

33. Avanzi C, del-Pozo J, Benjak A, Stevenson K, Simpson VR, Busso P, et al. Red squirrels in the British Isles are infected with leprosy bacilli. Science. 2016;354(6313):744.

34. Singh P, Benjak A, Schuenemann VJ, Herbig A, Avanzi C, Busso P, et al. Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis*. Proceedings of the National Academy of Sciences of the United States of America. 2015;112(14):4459-64.

35. Iwnetu R, van den Hombergh J, Woldeamanuel Y, Asfaw M, Gebrekirstos C, Negussie Y, et al. Is tuberculous lymphadenitis over-diagnosed in Ethiopia? Comparative performance of diagnostic tests for mycobacterial lymphadenitis in a high-burden country. Scandinavian journal of infectious diseases. 2009;41(6-7):462-8.

36. Lastoria JC, Abreu MA. Leprosy: review of the epidemiological, clinical, and etiopathogenic aspects Part 1. Anais brasileiros de dermatologia. 2014;89(2):205-18.

37. Thakkar S, Patel SV. Clinical profile of leprosy patients: a prospective study. Indian journal of dermatology. 2014;59(2):158-62.

38. Naveed T, Shaikh ZI, Anwar MI. Diagnostic Accuracy of Slit Skin Smear in Leprosy Pak Armed Forces Med J. 2015;65(5):649-52.

39. Martinez AN, Ribeiro-Alves M, Sarno EN, Moraes MO. Evaluation of qPCR-based assays for leprosy diagnosis directly in clinical specimens. PLoS neglected tropical diseases. 2011;5(10):e1354.

40. Lini N, Shankernarayan NP, Dharmalingam K. Quantitative real-time PCR analysis of *Mycobacterium leprae* DNA and mRNA in human biopsy material from leprosy and reactional cases. Journal of Medical Microbiology. 2009;58(6):753-9.

41. Azevedo MC, Ramuno NM, Fachin LR, Tassa M, Rosa PS, Belone AF, et al. qPCR detection of *Mycobacterium leprae* in biopsies and slit skin smear of different leprosy clinical forms. The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases. 2017;21(1):71-8.

Contributions

Design of the study: SG, AA, STC, CA, KD and KB

DNA extraction and PCR: SG, CA and PB

Staining and microscopic examination: SG

Histopathologic examination YT and MI

Identification of leprosy cases and patient management: SN

Data analysis: SG, STC, CA, TH, KD, KB and AA

Initial manuscript writing: SG, STC, CA, KD, KB and AA

Correction of the final manuscript: all authors

Funding information: this work was funded by the Swiss Coordination and Development Center (CODEV), the Fondation Raoul Follereau and the Swiss National Foundation grant IZRJ23-164174 and the Armauer Hansen Research Institute core fund.

S1 Table: Overview of the socio-demographic and clinical characteristic of leprosy cases included in this study with results of the laboratory tests – RR: reversal reaction, ENL: erythema Nosodum, N: neuritis, UCP: QIAmp UCP Pathogen kit, Fast DNA: QIAmp Fast DNA preparation kit

Project No.	Sex	Age	Number of lesions	Classificatio n	Clinical Diag	Family history	symtomm (Duratm)	Path H&E	HB/LB	Path. Classi	SSS-AFB	Fite Faraco	Gold Standard	Tissue-Auramine	SSS-Auramine	Extraction method	PCR	Disability	Reaction	Type of reacton	New,relaps	BI in skin lesions	BI in SSS
1	F	17	7	MB	POS	YES	8	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	HD	POS	NO	NO		NEW	0	0
2	M	28	2	PB	POS	YES	7	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	NO	NO		NEW	0	0
3	M	29	2	PB	POS	YES	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	YES	NO		NEW	0	0
4	M	65	9	MB	POS	NO	9	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	NO	YES	RR	NEW	0	0
5	M	30	8	MB	POS	NO	10	POS	LB	NEG	NEG	POS	POS	POS	NEG	HD	POS	YES	NO		NEW	1+	0
6	F	45	8	MB	POS	NO	15	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	NO	YES	RR	NEW	0	0
7	M	60	9	MB	POS	NO	11	POS	HB	BB	NEG	POS	POS	POS	NEG	HD	POS	NO	NO		NEW	4+	0
8	M	42	7	MB	POS	NO	12	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	3+	3+
9	F	35	12	MB	POS	NO	18	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	YES	YES	NU	NEW	0	0
10	F	30	10	MB	POS	NO	20	POS	HB	BL	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	6+	4+
11	F	55	10	MB	POS	NO	3	POS	HB	LL	POS	POS	POS	POS	POS	HD	POS	NO	YES	RR	NEW	6+	6+
12	M	66	16	MB	POS	NO	24	POS	HB	LL	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	6+	3+
13	M	30	13	MB	POS	NO	3	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	4+	4+
14	M	60	4	PB	POS	NO	8	POS	LB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	2+	4+
15	F	32	11	MB	POS	NO	17	POS	LB	NEG	NEG	POS	POS	POS	POS	HD	POS	YES	NO		NEW	1+	0
16	M	28	6	MB	POS	NO	11	POS	HB	NEG	NEG	NEG	POS	NEG	NEG	HD	POS	NO	NO		NEW	0	0
17	M	24	8	MB	POS	NO	12	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	YES	RR	NEW	2+	3+
18	M	34	6	MB	POS	NO	5	POS	LB	NEG	NEG	POS	POS	NEG	NEG	HD	POS	NO	NO		NEW	1+	0
19	M	45	11	MB	POS	NO	24	POS	HB	NEG	NEG	POS	POS	POS	NEG	HD	POS	YES	NO		NEW	2+	0
20	M	40	3	PB	POS	NO	16	NEG	LB	NEG	NEG	POS	POS	NEG	POS	HD	POS	YES	NO		NEW	1+	0
21	M	23	8	MB	POS	YES	24	POS	LB	NEG	NEG	POS	POS	POS	NEG	UCP	POS	NO	NO		NEW	3+	0
22	M	20	3	PB	POS	NO	20	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	2+	2+
23	M	15	14	MB	POS	YES	6	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	YES	N	NEW	4+	6+
24	M	30	2	PB	POS	NO	11	NEG	LB	NEG	POS	NEG	POS	NEG	NEG	UCP	POS	NO	NO		NEW	0	1+
25	M	40	3	PB	POS	NO	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	4+	4+
26	F	30	13	MB	POS	NO	24	POS	LB	NEG	NEG	NEG	POS	POS	NEG	UCP	POS	NO	NO		NEW	0	0
27	M	45	30	MB	POS	NO	8	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	YES	RR	NEW	5+	3+
28	F	35	8	MB	POS	NO	60	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	POS	YES	NO		NEW	0	0
29	M	30	20	MB	POS	YES	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	5+	6+
30	M	28	30	MB	POS	NO	6	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	4+	4+
31	M	57	30	MB	POS	YES	36	POS	LB	NEG	NEG	NEG	POS	POS	NEG	UCP	POS	YES	NO		NEW	0	0
32	M	21	20	MB	POS	NO	5	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	3+	6+
33	M	34	1	PB	POS	NO	5	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	YES	NO		NEW	0	0
34	F	30	20	MB	POS	NO	5	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	YES	YES	N	NEW	0	0
35	F	50	3	PB	POS	YES	NA	POS	HB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	YES	YES	RR	NEW	0	0
36	M	42	3	PB	POS	NO	2	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	NO	NO		NEW	0	0
37	M	35	30	MB	POS	NO	1	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		REL	6+	4+
38	M	67	20	MB	POS	NO	3	POS	HB	NEG	POS	POS	POS	POS	POS	UCP	POS	YES	NO		REL	1+	5+
39	M	27	10	MB	POS	NO	3	POS	HB	NEG	POS	POS	POS	POS	POS	UCP	POS	YES	YES	ENL	REL	1+	4+
40	F	28	20	MB	POS	YES	8	POS	HB	BL	POS	POS	POS	POS	POS	HD	POS	YES	YES	ENL	NEW	5+	5+
41	F	30	1	PB	POS	NO	48	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	YES	NO		NEW	0	0
42	F	55	10	MB	POS	NO	2	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	NO	YES	RR	NEW	0	0
44	M	40	20	MB	POS	NO	48	POS	HB	NEG	POS	POS	POS	POS	POS	UCP	POS	YES	YES	ENL	NEW	3+	3+
45	M	51	13	MB	POS	NO	1	POS	LB	NEG	NEG	NEG	POS	POS	NEG	UCP	POS	NO	YES	RR	NEW	0	0
46	M	42	13	MB	POS	NO		POS	LB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	YES	RR	NEW	4+	3+
47	M	28	12	MB	POS	NO	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	YES	YES	RR	NEW	0	0
48	M	38	20	MB	POS	NO	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	4+	3+
49	M	20	14	MB	POS	NO	24	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	6+	3+
50	M	15	16	MB	POS	YES	24	POS	HB	NEG	NEG	POS	POS	POS	POS	UCP	POS	YES	NO		NEW	1+	0
51	F	56	2	PB	POS	NO	20	NEG	LB	NEG	NEG	POS	POS	POS	POS	UCP	NEG	YES	NO		NEW	1+	0
52	M	60	10	MB	POS	NO	60	POS	HB	BL	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	5+	5+
53	M	35	2	PB	POS	NO	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	NO	NO		NEW	0	0
54	M	40	5	PB	POS	NO	4	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	NO	YES	RR	NEW	0	0
55	M	75	10	MB	POS	YES	6	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	UCP	NEG	YES	NO		NEW	0	0
56	M	60	6	MB	POS	NO	12	POS	LB	NEG	NEG	NEG	POS	POS	NEG	UCP	POS	YES	NO		NEW	0	0
57	F	27	30	MB	POS	NO	36	POS	HB	NEG	POS	POS	POS	POS	POS	UCP	POS	NO	NO		NEW	6+	3+
58	F	45	3	PB	POS	NO	24	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	YES	NO		NEW	0	0
59	M	47	11	MB	POS	NO	24	POS	HB	NEG	POS	NEG	POS	NEG	NEG	UCP	POS	YES	NO		NEW	0	3+

60	F	15	20	MB	POS	NO	12	POS	LB	NEG	NEG	POS	POS	POS	POS	UCP	POS	YES	NO		NEW	1+	0
61	M	30	10	MB	POS	YES	36	POS	LB	NEG	POS	POS	POS	POS	POS	UCP	POS	YES	YES	RR	NEW	1+	1+
62	M	26	20	MB	POS	NO	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	4+	3+
63	M	27	10	MB	POS	NO	12	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	6+	4+
64	F	50	2	PB	POS	YES	12	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	YES	NO		NEW	0	0
65	F	60	20	MB	POS	YES	NA	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	6+	4+
66	M	18	30	MB	POS	YES	36	POS	LB	NEG	POS	POS	POS	POS	POS	UCP	POS	YES	NO		NEW	3+	3+
67	F	55	20	MB	POS	NO	60	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	UCP	POS	NO	NO		NEW	0	0
68	M	60	30	MB	POS	NO	NA	POS	HB	NEG	POS	POS	POS	POS	POS	HD	POS	NO	NO		NEW	6+	3+
69	M	30	10	MB	POS	YES	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	NEG	NO	NO		NEW	0	0
70	F	15	10	MB	POS	YES	24	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	Fast DNA	POS	NO	NO		NEW	0	0
71	M	25	20	MB	POS	NO	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	UCP	POS	YES	NO		New	0	0
72	M	30	20	MB	POS	YES	24	POS	HB	BL	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	3+	2+
73	M	28	20	MB	POS	NO	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	4+	2+
74	M	25	9	MB	POS	YES	24	NEG	HB	NEG	POS	POS	POS	POS	POS	HD	POS	YES	NO		NEW	4+	2+
75	F	20	13	MB	POS	NO	10	POS	HB	NEG	NEG	POS	POS	POS	POS	UCP	POS	NO	NO		NEW	1+	0
76	F	38	1	PB	POS	NO	6	POS	LB	NEG	NEG	POS	POS	POS	POS	UCP	NEG	NO	NO		NEW	1+	0
77	F	20	20	MB	POS	YES	6	POS	HB	LL	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	4+	3+
78	M	21	10	MB	POS	NO	6	POS	LB	TT	NEG	POS	POS	POS	NEG	Fast DNA	NEG	YES	YES	RR	NEW	1+	0
79	F	28	10	MB	POS	NO	36	POS	LB	TT	NEG	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	1+	0
80	M	20	2	PB	POS	NO	96	POS	LB	TT	NEG	POS	POS	POS	NEG	Fast DNA	NEG	YES	NO		NEW	1scanty	0
81	F	43	2	PB	POS	NO	6	POS	HB	NEG	NEG	NEG	POS	NEG	NEG	Fast DNA	NEG	NO	NO		NEW	0	0
82	M	30	10	MB	POS	NO	12	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	YES	RR	NEW	3+	3+
83	F	30	2	PB	POS	NO	120	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	NEG	YES	NO		NEW	0	0
84	M	29	20	MB	POS	YES	6	NEG	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	YES	RR	NEW	3+	3+
85	F	68	6	MB	POS	NO	36	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	POS	YES	NO		NEW	0	0
86	F	15	110	MB	POS	NO	3	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	5+	5+
87	M	37	10	MB	POS	YES	36	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	Fast DNA	POS	YES	NO		NEW	0	0
88	F	48	3	PB	POS	NO	169	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	POS	YES	NO		NEW	0	0
89	M	30	20	MB	POS	YES	60	NEG	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	3+
90	M	35	10	MB	POS	NO	12	NEG	LB	NEG	NEG	POS	POS	POS	NEG	Fast DNA	POS	YES	NO		NEW	3+	0

91	M	32	14	MB	POS	YES	12	NEG	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	3+	2
92	F	60	20	MB	POS	NO	24	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	Fast DNA	POS	YES	NO		NEW	0	0
93	F	16	10	MB	POS	NO	12	NEG	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	6+	3
95	M	18	10	MB	POS	NO	6	NEG	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	3
96	M	22	10	MB	POS	YES	24	NEG	LB	NEG	POS	NEG	POS	NEG	POS	Fast DNA	POS	NO	NO		NEW	0	3
97	M	23	20	MB	POS	NO	36	POS	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	2+	1
98	M	70	40	MB	POS	NO	36	POS	LB	BT	NEG	POS	POS	POS	NEG	Fast DNA	POS	YES	NO		NEW	2+	0
99	F	25	15	MB	POS	NO	36	NEG	HB	LL	NEG	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	0
100	M	42	30	MB	POS	NO	48	POS	LB	BT	NEG	POS	POS	POS	POS	Fast DNA	POS	YES	YES	RR	NEW	2+	0
101	F	40	20	MB	POS	NO	36	NEG	LB	NEG	NEG	POS	POS	POS	NEG	Fast DNA	POS	YES	NO		NEW	1+	0
103	M	37	4	PB	POS	NO	120	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	Fast DNA	POS	YES	NO		NEW	0	0
104	M	26	20	MB	POS	NO	12	NEG	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	3+	3
105	F	45	15	MB	POS	YES	12	POS	LB	NEG	NEG	POS	POS	POS	NEG	Fast DNA	POS	NO	NO		NEW	1+	0
106	F	45	14	MB	POS	NO	12	POS	HB	LL	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	3+	3
107	M	35	6	MB	POS	YES	24	NEG	LB	NEG	NEG	POS	POS	POS	NEG	Fast DNA	POS	NO	NO		NEW	1+	0
109	M	40	12	MB	POS	NO	60	POS	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	YES	RR	NEW	3+	2
110	M	53	50	MB	POS	NO	12	NEG	HB	LL	NEG	POS	POS	POS	NEG	Fast DNA	POS	NO	NO		REL	6+	0
111	M	21	10	MB	POS	YES	12	POS	HB	BB	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	5
112	M	20	15	MB	POS	NO		POS	LB	IND	NEG	POS	POS	NEG	NEG	Fast DNA	NEG	YES	YES	RR	NEW	1+	0
113	M	29	4	PB	POS	YES	1	NEG	LB	NEG	POS	NEG	POS	NEG	NEG	Fast DNA	POS	YES	NO		NEW	0	2
114	M	29	20	MB	POS	NO	12	POS	LB	NEG	NEG	NEG	POS	POS	NEG	Fast DNA	POS	NO	YES	RR	NEW	0	0
115	M	30	10	MB	POS	NO	12	POS	LB	NEG	POS	NEG	POS	NEG	POS	Fast DNA	POS	YES	NO		REL	0	2
116	M	28	5	PB	POS	NO	2	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	POS	YES	NO		NEW	0	0
117	F	20	30	MB	POS	NO	6	POS	HB	BB	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	5
118	F	28	20	MB	POS	NO	84	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	YES	ENL	NEW	6+	6
119	M	32	4	PB	POS	YES	4	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	NEG	YES	NO	N	NEW	0	0
120	M	23	10	MB	POS	NO	1	POS	HB	LL	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	4+	6
121	M	31	20	MB	POS	NO	72	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	6
122	M	50	10	MB	POS	NO	12	NEG	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	6+	6
123	F	23	10	MB	POS	YES	12	POS	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	4+	4
124	F	65	10	MB	POS	YES	8	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	6

126	F	35	10	MB	POS	NO	18	NEG	NA	NEG	NEG	NEG	NEG	NEG	NEG	Fast DNA	NEG	YES	YES	N	NEW	0	0
127	M	50	10	MB	POS	NO	24	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	6+
128	M	35	10	MB	POS	NO	36	POS	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	6+	6+
129	M	18	20	MB	POS	NO		NEG	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	4+	4+
130	M	25	19	MB	POS	NO	24	NEG	HB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	5+	4+
131	M	25	1	PB	POS	NO	72	NEG	NA	NEG	NEG	NEG	NEG	POS	NEG	Fast DNA	NEG	NO	NO		NEW	0	0
132	M	22	3	PB	POS	NO	24	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	Fast DNA	NEG	NO	NO		NEW	0	0
133	M	28	32	MB	POS	NO	24	POS	LB	IND	NEG	NEG	POS	NEG	NEG	Fast DNA	POS	YES	NO		NEW	0	0
134	M	35	20	MB	POS	NO	8	POS	HB	BL	POS	POS	POS	POS	POS	Fast DNA	POS	YES	NO		NEW	4+	3+
135	M	65	10	MB	POS	NO	24	POS	HB	BL	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	3+	2+
136	M	28	10	MB	POS	YES	0	POS	LB	IND	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	45	5+
138	M	35	10	MB	POS	NO	24	POS	HB	LL	POS	POS	POS	POS	POS	Fast DNA	POS	NO	NO		NEW	4+	5+
139	F	20	20	MB	POS	NO	8	POS	LB	NEG	NEG	NEG	POS	NEG	NEG	Fast DNA	NEG	YES	YES	RR	NEW	0	0
140	M	35	10	MB	POS	NO	12	POS	LB	BT	NEG	NEG	POS	NEG	NEG	Fast DNA	NEG	YES	YES	RR	NEW	0	0
141	M	43	24	MB	POS	NO	102	NEG	LB	NEG	NEG	POS	POS	NEG	NEG	Fast DNA	NEG	NO	NO		NEW	1+	0
142	M	22	30	MB	POS	NO	48	POS	LB	NEG	POS	POS	POS	POS	POS	Fast DNA	POS	YES	YES	ENL	NEW	1+	4+

S2 Table: Overview of the socio-demographic and clinical characteristic of non-leprosy cases included in this study with results of the laboratory tests and other diseases associated

Project No.	Sex	Age	Clinical diag	Family history	Path H&E	Gold standard	Tissue-Auramine	Fite Faraco	PCR	Case for surgery
ARLP 143	F	65	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 144	M	19	NEG	NO	NEG	NEG	NEG	NEG	NEG	excisional
ARLP 145	M	55	NEG	NO	NEG	NEG	NEG	NEG	NEG	Incsional
ARLP 146	F	22	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 147	M	60	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 148	M	60	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 149	M	30	NEG	NO	NEG	NEG	NEG	NEG	NEG	Lymphoma
ARLP 150	F	50	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 151	F	58	NEG	NO	NEG	NEG	NEG	NEG	NEG	Lipoma
ARLP 152	M	28	NEG	NO	NEG	NEG	NEG	NEG	NEG	carcinosis
ARLP 153	F	27	NEG	NO	NEG	NEG	NEG	NEG	NEG	Correction abputation
ARLP 154	M	54	NEG	NO	NEG	NEG	NEG	NEG	NEG	Carcinosis
ARLP 155	F	40	NEG	NO	NEG	NEG	NEG	NEG	NEG	Flap Division
ARLP 156	F	24	NEG	NO	NEG	NEG	NEG	NEG	NEG	Wound closure
ARLP 157	F	70	NEG	NO	NEG	NEG	NEG	NEG	NEG	correction aputation
ARLP 158	F	58	NEG	NO	NEG	NEG	NEG	NEG	NEG	Correction ambutation
ARLP 159	M	25	NEG	NO	NEG	NEG	NEG	NEG	NEG	keloid
ARLP 160	F	23	NEG	NO	NEG	NEG	NEG	NEG	NEG	keloid
ARLP 161	F	25	NEG	NO	NEG	NEG	NEG	NEG	NEG	scare
ARLP 162	F	65	NEG	NO	NEG	NEG	NEG	NEG	NEG	excisional
ARLP 163	F	20	NEG	NO	NEG	NEG	NEG	NEG	NEG	keloid
ARLP 164	F	21	NEG	NO	NEG	NEG	NEG	NEG	NEG	keloid
ARLP 165	F	24	NEG	NO	NEG	NEG	NEG	NEG	NEG	keloid
ARLP 166	M	20	NEG	NO	NEG	NEG	NEG	NEG	NEG	corn
ARLP 177	F	23	NEG	NO	NEG	NEG	NEG	NEG	NEG	Scare
ARLP 168	F	32	NEG	NO	NEG	NEG	NEG	NEG	NEG	Keratin
ARLP 169	F	27	NEG	NO	NEG	NEG	NEG	NEG	NEG	Lymphoma
ARLP 170	F	28	NEG	NO	NEG	NEG	NEG	NEG	NEG	cyst

S3 Table: Histopathological repartition of the gold standard positive and negative samples with the number of positives and positivity rate (%) for each laboratory tests – HB: high BI included samples from the LL, BL and BB groups; LB: low BI included samples from the BT, TT, IND and NEG groups – The table show the high positivity rate for AO in punch biopsies and PCR compared to other methods.

		Number of sample	SSS		Punch biopsy		
			ZN(%)	AO (%)	FF (%)	AO (%)	PCR (%)
Gold standard positive	LL	24	22(91.7)	23(95.8)	24(100)	24(100)	24(100)
	BL	20	19(95)	19(95)	20(100)	20(100)	20(100)
	BB	14	8(57.1)	9(64.3)	10(71.4)	10(71.4)	13(92.9)
	BT	26	7(26.9)	12(46.1)	14(53.8)	17(65.4)	24(92.3)
	TT	9	0(0)	2(22.2)	4(44.4)	5(55.6	5(55.5)
	INT	6	3(50)	3(50)	4(66.7)	3(50)	5(83.3)
	NEG	14	8(57.1)	8(57.1)	11(78.6)	9(64.3)	12(85.7)
	Total	113	67(59.3)	76(67.3)	87(77)	88(77.9)	103(91.1)
	HBL	58	49(84.5)	51(87.9)	54(93.1)	54(93.1)	57(98.2)
	LBL	55	18(32.7)	25(45.4)	33(60)	34(61.8)	46(83.6)
Gold standard negative		24	0(0)	0(0)	0(0)	6(25)	10(41.7)

S4 Table: Statistical value obtained with the binomial tests and the Fisher test for the different tests in different condition (all samples, LB samples or HB samples)– p= p-value in red when p> 0.05 and in green when *p*<0.05; OR: odds ratio;

		All samples					LB samples					
		ZN	AO in SSS	FF	AO in tissue	PCR_{m+c}	ZN	AO in SSS	FF	AO in tissue	PCR	PCR_{m+c}
All samples	AO in SSS	*p*=0.09										
	FF	*p*=0.006 OR= 2.3	*p*=0.08 OR= 1.75									
	AO in tissue	*p*=0.003 OR= 2.4	*p*=0.05 OR= 1.85	*p*=1								
	PCR			*p*=0.005 OR=3	*p*=0.009 – OR=2.9							
	PCR_c					*p*=0.04 OR= 5.6						
HB	ZN						*p*=1.7e-8 OR=10.9					
	AO in SSS							*p*=3.5e-7 OR=9.9				
	FF								*p*=3.3e-5 OR=8.8			
	AO in tissue									*p*=7.2e-5 OR=8.2		
	PCR										*p*=0.007OR=10.9	
LB	AO in SSS						*p*=0.2					
	FF						*p*=0.0004	*p*=0.01				
	AO in tissue						*p*=0.01	*p*=0.01	*p*=1			
	PCR						*p*=2.2e-16	*p*=9.8e-9	*p*=0.01	*p*=0.01		
	PCR_{m+c}											
	PCR_c											*p*=0.15, OR= 3.7

S1 Fig: Raw data used for the calculation of the sensivity, specificity, PPV and NPV of the routine methods, AO and PCREach value represented the number of patients per group

ZN	Gold standard +	Gold standard -
+	67	0
-	46	24

FF	Gold standard +	Gold standard -
+	87	0
-	26	28

AO (SSS)	Gold standard +	Gold standard -
+	74	0
-	39	24

AO (biopsy)	Gold standard +	Gold standard -
+	88	0
-	25	28

PCR	Gold standard +	Gold standard -
+	103	0
-	10	28

Chapter 7: Conclusions and perspectives

The rise of genomics has opened an unprecedented, accessible and faster way to obtain deeper insight into the evolution, transmission, pathogenicity and resistance mechanisms of infectious disease, including leprosy bacilli. At the beginning of the 21st century, whole-genome sequencing rapidly overcomes standard PCR-sequencing techniques to study infectious diseases. However, PCR-sequencing is still predominately used in the leprosy field. One of the main reasons for this is the high cost of the whole-genome sequencing technique applied to host-derived tissues, in contrast to pure bacterial cultures.

Since the sequencing of the first *M. leprae* genome in 2001 (1), Prof. Cole's laboratory has dedicated part of its research to the development of tools to monitor drug resistance and to understand leprosy transmission using genomics approaches and especially whole-genome sequencing. This work led to significant breakthroughs in the leprosy field and the associated development of robust research tools (2–6).

In this context, my role was to develop simpler and more efficient tools to facilitate *M. leprae* genome sequencing. I conducted my doctorate programme in a multidisciplinary team of researchers including molecular bacteriologists, biochemists, bioinformaticians, and skilled technicians. Moreover, since leprosy is not endemic in Switzerland, I collaborated with many researchers throughout the world (Acknowledgments map p.229), which immersed me in the field and helped me to understand all the challenges of leprosy research. In these collaborative projects, the main goal was to develop and apply whole-genome sequencing tools in order to solve specific challenges around leprosy. Thus, in this final part of the thesis, I would like to summarize the main findings and discuss further perspectives and future work.

M. lepromatosis: the new but still mysterious agent of leprosy

The whole genome sequencing of *M. lepromatosis* revealed that the new leprosy agent and *M. leprae* are distinct but closely related species (Fig.1). They share 92% nucleotide identity, far less compared to the bovine tuberculosis agent *M. bovis* and the human-associated agent *M. tuberculosis* (99.95%) (9). Interestingly, they both share the same level of extensive pseudogenization, probably linked with the intracellular adaptation of their most recent common ancestor. More strikingly, despite their divergence, clinical outcomes are similar, and it is not possible to distinguish between *M. leprae* and *M. lepromatosis* infections without additional laboratory tests.

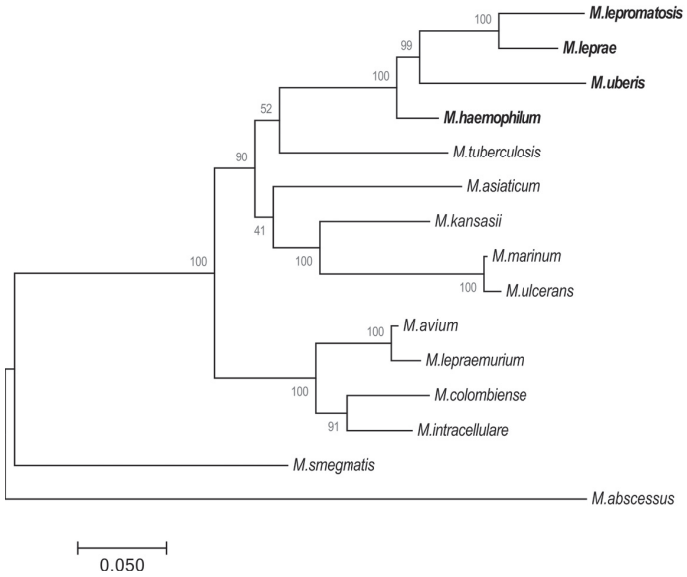


Figure 1: Maximum likelihood tree of *M. leprae* and selected mycobacterial species. The tree was created using MEGA7 from concatenated amino acid sequences of 11 proteins (DnaN, RpII, GrpE, MetG, RplY, PheT, FtsQ, HolA, MiaA, FtsY and, FtsX). (7,8)

Using genomics, it is now possible to target unique genes of *M. lepromatosis* (such as the *hemN* gene), absent in the *M. leprae* genome, allowing specific differentiation of these species. Repetitive regions are other sensitive targets to consider for future diagnosis purposes. Unfortunately, most of these regions are not covered in the current *M. lepromatosis* assembly because it was based on short Illumina reads. Prospective efforts to sequence these repetitive regions using long read sequencing are indispensable for developing sensitive and specific diagnostic tests as exemplified in *M. leprae* with the specific RLEP sequences (10).

M. lepromatosis infection is almost exclusively reported in Mexico and the Caribbean region, with a few sporadic cases in Asia and North America (11–14). A survey of 120 patients in Mexico, with various clinical forms of leprosy, revealed that *M. lepromatosis* infection occurs more often than *M. leprae* in DLL cases (63.2%) and that some cases (16.1%) might be dual infections (15). In our screen of leprosy samples from Venezuela, Mali, Brazil, and Mexico, *M. lepromatosis* was identified only in Mexico, and was associated with lepromatous leprosy, DLL and Lucio's phenomenon.

Strikingly, DLL and Lucio's phenomenon are also historically principally found in Mexico and Central America (16,17) where *M. lepromatosis* is mostly identified. However, cases of Lucio's phenomenon were also reported from other regions such as in Sri Lanka (18), Perù (19), India (20–24), Brazil (25–27), Malaysia (28), USA (11,29), Singapore (12), Canada (13), Myanmar (14) Tunisia (30) and Iran (31). Unfortunately, the molecular characterization of the causative agent is rarely performed.

DLL associated with Lucio's phenomenon is rare but fatal without early treatment. The spark that triggers this specific leprosy form is unknown. The reason could be an environmental factor or the host immune response but also the genetics of the infectious agent. Systematic sequencing of the leprosy bacilli in case of DLL/Lucio's phenomenon and comparative genomics, combined with a transcriptomic analysis of the host and pathogen in comparison with other forms of leprosy, could help to identify the aetiology of the DLL form.

The only non-human *M. lepromatosis* reservoir known to date is the red squirrel in the British Isles. When present, symptoms are similar to human leprosy, including nerve involvement, and tropism for the cooler parts of the body such as the muzzle, ears, and limbs. Genomic comparison between the Mexican and the British strains revealed extreme sequence conservation of the strains from different areas. However, it is unclear whether the British strain could infect humans similar to the Mexican one because of the differences in the genome. Also, the origin of the infection in red squirrels is mysterious since, as yet, there is no evidence of human infections in the past. Therefore, analysis of ancient DNA from human remains might help to retrace the origin and the spread of this pathogen.

Taking the data together, one could be tempted to make the parallel with the current epidemiology and history of *M. bovis*. As part of the *M. tuberculosis* complex, *M. bovis* is the cause of bovine tuberculosis in cattle and in many animals including man and non-human primates (32). These infections occurred in small pockets at different levels depending on the animal infected (32,33). Moreover, *M. bovis* tuberculosis is clinically indistinguishable from *M. tuberculosis* infection in humans similar to *M. leprae* and *M. lepromatosis* infections. The main source of human infection of *M. bovis* is cattle because humans consume cattle products such as meat and milk. Also, in countries where tuberculosis in cattle is common, such as Mexico, up to 14% of the human cases are caused by *M. bovis* while the WHO estimated that *M. bovis* is responsible for 3,1% of all tuberculosis cases in humans worldwide (34). However, for other animals such as deer, possum, ferret, pigs or llama the animal-to-human transmission is more limited (35,36). For leprosy, one hypothesis would be that *M. lepromatosis* is an animal adapted leprosy bacillus that can eventually be transmitted to humans in contact with infected animals. This could explain why *M. lepromatosis* is identified to a lesser extent compared to *M. leprae* in humans. In parallel, little is known about the global burden of *M. lepromatosis* infection nowadays, and new epidemiological surveys with the systematic sequencing of the strains should be conducted in Mexico, the surrounding countries and for each suspected

leprosy case where *M. leprae* is not found to identify the transmission dynamics of the bacteria.

Towards an *M. leprae* complex?

In 2017, the concept of an *M. leprae* complex has emerged following the report of two leprosy-like diseases in wild animals where the causative agents shared genetic similarities with the leprosy bacilli (37). One of the new mycobacterial species was identified in cows in France as causing bovine nodular thelitis (38). Sequencing of six genes revealed that this new species is related to *M. leprae* and *M. lepromatosis* (38). Recently, whole genome sequencing of this new species, named *M. uberis*, was achieved in our laboratory after DNA extraction from cow tissue (data not shown). The genome was assembled in 56 contigs amounting to 3.1 Mb. Preliminary analysis confirmed the close relatedness of *M. uberis* with the leprosy bacilli (Fig. 1). Also, *M. uberis* underwent genome downsizing, often associated with extensive reductive evolution. Further analysis of the pseudogenization rate is ongoing.

The second species, named *M. lepraefelis*, causes feline leprosy in domestic cats in Australia and New Zealand (39). Preliminary sequencing also revealed a close genetic relationship with the leprosy bacillus, but the whole genome sequence is not available yet.

The parameters to define a complex are not described in the literature, but several examples shared the clinical outcome as the main feature independently or not of genetics (40,41). Both new mycobacterial species have a tropism for cool parts of the body and are not cultivable *in vitro*, the same as the leprosy bacillus (38,39). Histopathological features are similar to leprosy disease; however, bacilli are not observed in nerves, a specificity of leprosy. Besides, the clinical outcome is identical for *M. lepraefelis* and the leprosy bacillus [except for the peripheral neuropathy] with disseminated skin lesions whereas *M. uberis* seems to lead to lesions on the cow teat. Finally, the two new species are not associated with human cases so far. Thus, *M. leprae* and *M. lepromatosis* could be classified in the same complex whereas *M. uberis* and *M. lepraefelis* would be part of another complex based on the clinical outcome.

Nevertheless, these organisms, regardless of the absence of nerve involvement, might have some common biological aspects, like their cell infection mechanism or interaction with the host. Studying them might improve our understanding of the biology of the leprosy bacilli.

Phylogeny of *M. leprae* strains

One of the main objectives of my thesis was to increase the number of sequenced genomes from different parts of the word in order to improve the resolution of the *M. leprae* phylogeny and to identify the geographic origin of the bacterium. Since 2013, we have increased the number of sequenced *M. leprae* genomes by 20-fold, with 283 genomes available to date (Fig 2). This dataset gives an unprecedented overview of the current phylogeny of *M. leprae*, but also raises new questions and hypotheses.

The current tree conformation fits well with the previous SNP typing code established by Monot *et al*. (4). Unfortunately, we did not obtain whole genomes for the SNP subtypes 1C, 3G et 3I to confirm their classification. Also, new branches were identified such as the 3K1, discussed below.

Monot *et al*. suggested that leprosy originated in East Africa or the Near East with SNP type 2 or SNP type 3 *M. leprae* and that the ancestral strain evolved and propagated via past human migrations (4). We confirmed that the SNP type 2 is not the ancestral lineage, nor is SNP type 3M. In parallel, our results corroborated the observation of Schuenemann *et al*. that the SNP subtype 3K0 (Branch 0) is the ancestral lineage (6). The propagation of this lineage was probably boosted by the Silk Roads, but the starting point remains unclear. India is surrounded by both Silk Roads and is today the country with the highest number of leprosy cases. India was previously hypothesized to be the origin of leprosy in Europe when Alexander the Great's troops came back from the Indian campaign (42). SNP types 1 and 2 are phylogenetically recent lineages and currently predomine in India, suggesting that the country was not the starting point of leprosy for Europe, which contained almost all lineages of *M. leprae*, including the ancestral 3K (43). We discovered a new lineage, named 3K1, from a

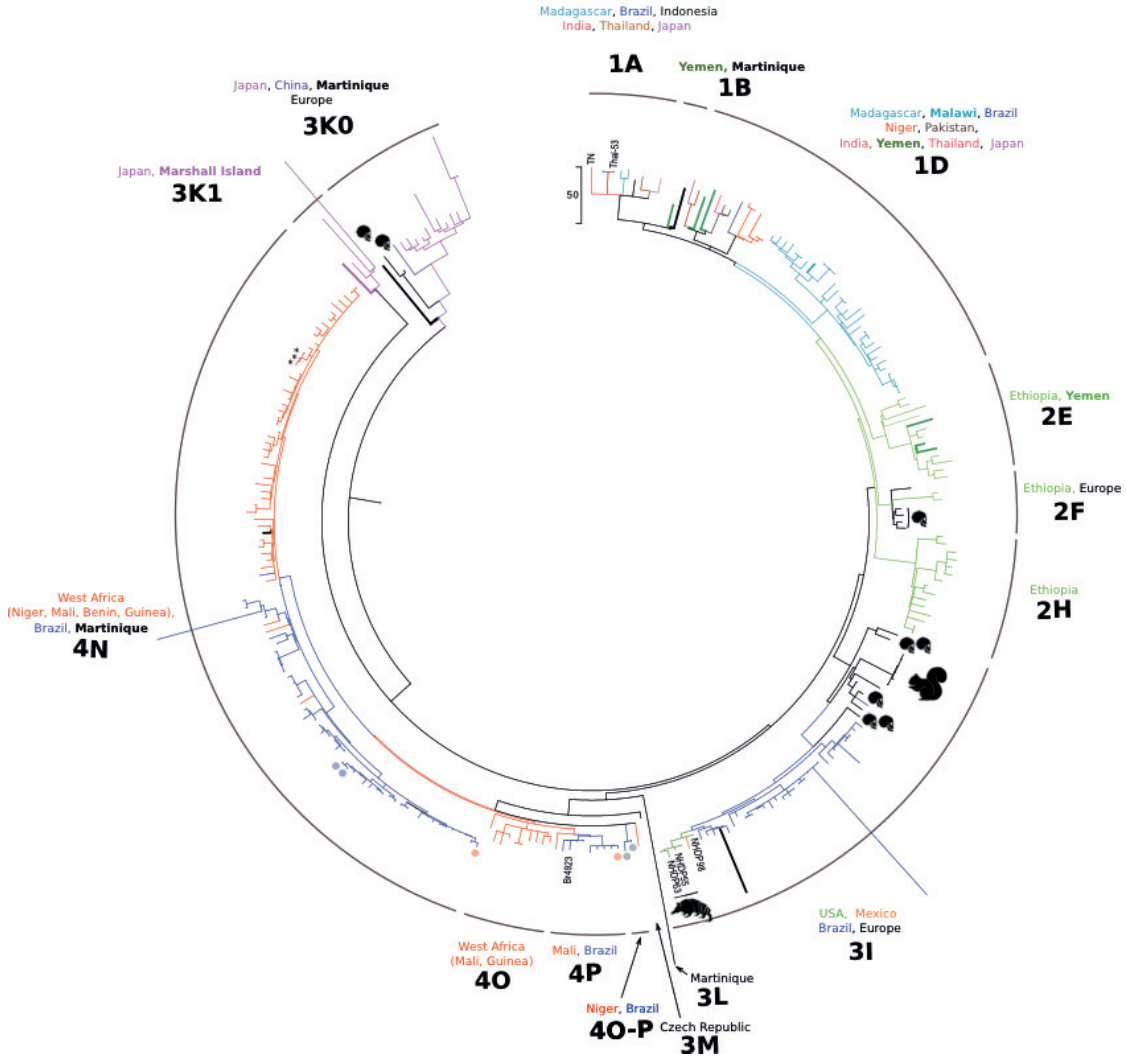


Figure 2: Maximum parsimony tree of 279 medieval (n=15) and modern *M. leprae* strains (n=264), rooted with *M. lepromatosis* – The different geographical origin and SNP subtypes given for each SNP –subtype. Samples from the relapse and reinfection study are indicated with colored squares (blue, red and green) while samples from Guinea-Conakry with the specific *folP1* mutation are indicated with the *. Medieval samples are represented by skulls.

few genomes from Japan and the Marshall Islands.

A PCR screen of additional samples performed in our laboratory (data not shown) also revealed its presence in the Philippines, Korea, and Turkey. Interestingly, the SNP type 3K seems to be widely spread in China with the presence of the 3K0 in the West part of the country (44), but it has never been reported in India despite its proximity to China (44,45). One of the explanations for the absence of the 3K genotype in India could be the geographic and cultural barrier between the two countries, resulting in limited migrations of peoples. Thus, the 3K lineage was probably distributed along inland Asia. However, since the number of available genomes from India is too low to represent the entire Indian population, we cannot exclude the possibility of the existence of a more ancestral unknown lineage. Nevertheless, an exhaustive sequencing effort from this part of the world would provide more conclusive evidence of the past spread of leprosy. Similarly, limited information is available for other countries along the Silk Road, especially Central Asia. Monot *et al*. reported the SNP type 3K in Iran and Turkey (4), but no sequencing was performed from these countries. Obtaining samples from this part of the world is particularly difficult due to geopolitical reasons. Moreover, China, the biggest country in Asia, and adjacent countries associated with the Silk Roads (Myanmar, Vietnam, Laos, Mongolia, Nepal, Pakistan) should be extensively covered to get more resolution and potential insight into diversity of the ancestral 3K lineage.

During my thesis, efforts were also made to sequence strains from Brazil, the second country, after India, with the highest number of leprosy cases worldwide (46). Almost 100 genomes were sequenced showing great genomic diversity, including SNP subtype 3I, 4N, 4P, and 1D, which probably resulted from multiple introductions from different parts of the word during the post-Colombian colonization of South America. Interestingly, some of the SNP subtype 3I strains from Brazil belong to the most ancestral medieval European lineages (manuscript in preparation). Leprosy was, in part, brought to the Americas by European settlers, and Brazil might still carry the genomic diversity of *M. leprae* from the post-Colombian Europe.

In the *M. tuberculosis* complex, the Beijing lineage is thought to be the most successful lineage, with increased virulence and rapid acquisition of drug resistance (47). These faculties are explained by a cumulative role of environmental factors, immune host pressure and strain genetics (47). For *M. leprae*, several SNP types are rare nowadays such as 3M, 3J, 1B, 1C, 2G and the new 4O/P lineage. On the other hand, SNP subtype 3I seems to be spreading more successfully, including to non-human hosts. SNP type 3K is also widespread, despite its age. It is tempting to speculate that different *M. leprae* lineages adapted to different host genetics, or became more virulent or contagious. Virulence and transmission rates are difficult to measure for leprosy because of the wide range of symptoms, long incubation times and slow progress of the disease. However, a study in Columbia showed an association of European (SNP type 3) and African (SNP type 4) *M. leprae* strains with the ancestral origin of the patient (from Europe or Africa) (48). Therefore, the rarity of some SNP types might be due to strain replacement or the geographic isolation of some populations. However, these variations in strain distribution could also suggest less successful adaptation to the local environment or host population. Studying a large cohort in a multi-ethnic locale, such as Brazil, for example, could be a good approach to investigate *M. leprae*'s adaptation to its host or the role of social factors like population density or sanitation. Finally, another important aspect is the possible link between the genetics of the *M. leprae* strain and disease outcome. So far, the aetiology of the wide spectrum of leprosy is explained by the immune status of the patient at the time of infection. However, a link between a particular polymorphism (SNP subtype or an antigen variation) and a specific clinical outcome should not be ruled out. This analysis could be achieved in the next few years with increased whole genome sequencing of strains from well-characterized patients.

Tracking leprosy transmission

Variation of M. leprae strains

M. leprae is a clonal organism with a very low substitution rate, leading to extremely limited genetic variation between strains from

different areas but also within the same strain from different time points.

This limited variation was first observed in two siblings from Guinea-Conakry who both developed leprosy within a few years. The isolates differed by only two polymorphisms (Fig. 2). An isolate from an unrelated patient from the same province showed only one additional mutation. Similarly, we recently sequenced a strain isolated at different treatment time points and from different skin lesions of the same patient from Madagascar and observed zero, or only one polymorphism between the isolates (manuscript in preparation). A similar low variation was also observed in the red squirrel population in Brownsea Island.

The low substitution rate of *M. leprae* poses limitations for short-range transmission studies, including characterization of relapse cases. For example, among the three recurrent Brazilian cases analysed in 2016 (Fig. 2), one was undoubtedly reinfected with a different strain four years after the first episode, but for the two other cases, isolates spanning seven and eight years differed only two polymorphisms with no mutation in the known drug-resistance determining regions. Such cases remain inconclusive because they could be re-infections with a similar strain circulating in the area or true relapses with the original strain.

The mechanism of relapse in the absence of drug resistance is unknown in leprosy. Apart from treatment failure and non-adherence to the treatment, the existence of more virulent, or persistent strains could explain this phenomenon, similar to *M. tuberculosis* (49–53). Systematic and large-scale studies based on whole genome sequencing of isolates from relapse cases will help to identify such mechanisms.

Another interesting question is the possibility of a specific strain triggering protection after primary infection assuming that this protection would only be conferred for a specific strain and for a limited period of time as is the case for other pathogenic bacteria such as *Streptococcus pneumoniae* (54). This is an attractive hypothesis suggesting that reinfection could only occur with different strains. However, identification of such markers will be challenging and will require advanced –omics methods of investigation such as functional antibody repertoire sequencing (55).

The red squirrel reservoir

Until 2016, the nine-banded armadillo was the only known animal reservoir of leprosy bacilli. The discovery of *M. leprae* in the red squirrel population on Brownsea Island was unexpected in a country where leprosy is no longer identified in autochthonous human cases. The strain is closely related to the one circulating in medieval Europe, suggesting that the squirrel might have been infected during this period when leprosy was at its peak in Europe. We were not able to identify *M. leprae* in other parts of the British Isles where the population of red squirrels is still present. It is possible that the bacteria have survived on Brownsea Island because the reservoir is isolated and protected, and the high density of squirrels on the island probably favors the transmission of the pathogen. The repercussion of this discovery is important because, if an animal reservoir can survive long after the disappearance of the disease in humans, important reservoirs might exist in endemic countries that could contribute to the continuous spread of the disease. Based on the host preferences of other mycobacteria and the known hosts for the leprosy bacilli, potential new reservoirs likely include other mammals, although other animals cannot be excluded, either as hosts or vectors.

Drug resistance and compensatory mutations

There are six drugs available to treat leprosy, three of which have an unknown mechanism of action. Clofazimine's mechanism is especially important to elucidate because it is also used to treat multi-drug resistant *M. tuberculosis* strains (56). In addition, as exemplified by the rifampicin-resistant Airaku-3 strain, which has a wild-type *rpoB*, unknown mechanisms of resistance still await elucidation.

Standard susceptibility tests cannot be easily performed for *M. leprae*. Thus, we overtook the challenge of genome sequencing directly from host tissues and obtained over twenty genomes from drug-resistant strains. This allowed us to identify probable

compensatory mutations in *rpoB*, *rpoC*, and *rpoA* that need to be functionally validated. During our analysis, we identified a couple of highly mutated genes associated with drug-resistant strains, such as *ribD, fadD9* and *nth* coding for a dihydropteroate synthase analog, an acyl-CoA synthetase and the endonuclease III respectively. While the impact of the *nth* polymorphism was easily associated with a hypermutator phenotype, the challenge remains to confirm the role of the other mutations. Empirically, the MFP assay is the reference method for drug susceptibility test, but it is inappropriate for a large screening.

Current *in vitro* tests include gene replacement in *Escherichia coli* by specific *M. leprae* genes. This method was successfully applied to describe the drug resistance mechanism of *M. leprae* to dapsone (57). However, *E. coli* is a Gram negative bacterium and shows different drug susceptibility patterns compared to Gram-positive bacteria (58). Another approach is to use close relatives such as the non-pathogenic *M. smegmatis*, or *M. tuberculosis*. Studies can be conducted using gene replacement, or by direct mutagenesis of the genome (59–61). However, *M. smegmatis* and *M. tuberculosis* harbor four and six times more coding genes, respectively, compared to *M. leprae* and the alternative pathways, or genes with redundant functions, might mask the real effect of the targeted mutation.

Phylogenetically, *M. haemophilum* is the closest cultivable species to the leprosy bacillus. It grows at a low temperature (30°C to 32°C), and it requires iron supplementation, such as hemin for *in vitro* growth, but it was shown that it could be engineered to express foreign genes. Thus, *M. haemophilum* could be an alternative candidate to elucidate the function of *M. leprae* genetic variants (62).

Adaptation and persistence

In addition to of the known drug-resistance associated genes, other hypermutated genes were identified in *M. leprae* without an obvious link to drug-resistance, such as some PPE and PE proteins, a putative nucleotide cyclase, a probable transcriptional regulatory protein, a putative ribonuclease J, and some genes with unknown function. Some of these genes, like *ML0411,* have no orthologs in other mycobacteria, while the function of the other genes is not well described in mycobacteria, making it difficult to link them with particular phenotypes. However, for a species with a reduced number of coding genes, such a high rate of mutations probably led to important biological functions such as persistence, dormancy or virulence.

Diagnosis of leprosy

Efficient diagnosis of leprosy during the early stages of the disease and identification of the different clinical forms is one of the biggest challenges of leprosy control, especially for the cases with no obvious clinical signs. During my thesis, I worked on two projects tackling the problem from different angles. The first project aimed to compare the whole transcriptome of leprosy patients with healthy endemic controls in order to identify specific signatures of the disease in the blood of patients. This work is still in progress. The second project was to improve the direct detection of the leprosy bacilli in the skin using molecular methods, which can be more sensitive than standard microscopy. However, even molecular methods, such as PCR, do not perform well for samples with low BI. To overcome the problem, I first improved the most critical step, DNA extraction from punch biopsies. Indeed, efficient lysis of the bacteria is crucial to maximise the recovery of DNA. When chemical lysis is complemented with mechanical lysis, the sensitivity of the method improved, especially for samples with a low BI.

The second step where optimization is required is the detection. The main drawbacks of conventional PCR are the risk of amplicon contamination and additional detection steps are needed, like gel electrophoresis, compared to quantitative PCR where amplified DNA is contained, and results obtained are in real time. During the last few months of my thesis, I optimized a qPCR method combined with efficient DNA extraction in order to the improve detection rate in low BI samples. The validation of the method is in process.

Overall, our results show that with an efficient extraction procedure, enough bacterial DNA is recovered to allow sensitive detection of the pathogen. This method can be applied to even more challenging samples

such as blood, slit skin smears, urine, nasal swab or fine needle aspirates.

Conclusion

The results obtained in my thesis work have radically changed our understanding of the origin and phylogeny of *M. leprae* and established its relationship to other mycobacterial pathogen like *M. lepromatosis.* Our results also highlight the need for systematic sequencing of all *M. leprae* isolates, especially from drug-resistant or recurrent cases, in order to identify new biological mechanisms. In parallel, efforts should be made to develop tools to automate processing and analysis of data, and new models should be developed that allow functional studies *in vitro*. Leprosy is a challenging disease to study because most of the standard microbiological tools cannot be used. This is where -omics methods can overcome this challenge and help fight this devastating disease.

References

1. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, et al. Massive gene decay in the leprosy bacillus. Nature. 2001 Feb 22;409(6823):1007–11.

2. Honore N, Cole ST. Molecular basis of rifampin resistance in *Mycobacterium leprae*. Antimicrob Agents Chemother. 1993 Mar;37(3):414–8.

3. Monot M, Honoré N, Garnier T, Araoz R, Coppée J-Y, Lacroix C, et al. On the origin of leprosy. Science. 2005 May 13;308(5724):1040–2.

4. Monot M, Honoré N, Garnier T, Zidane N, Sherafi D, Paniz-Mondolfi A, et al. Comparative genomic and phylogeographic analysis of *Mycobacterium leprae*. Nat Genet. 2009 Dec;41(12):1282–9.

5. Truman RW, Singh P, Sharma R, Busso P, Rougemont J, Paniz-Mondolfi A, et al. Probable zoonotic leprosy in the southern United States. N Engl J Med. 2011 Apr 28;364(17):1626–33.

6. Schuenemann VJ, Singh P, Mendum TA, Krause-Kyora B, Jäger G, Bos KI, et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science. 2013 Jul 12;341(6142):179–83.

7. Mizuno T, Natori T, Kanazawa I, Eldesouky I, Fukunaga T, Ezaki T. Core housekeeping proteins useful for identification and classification of mycobacteria. Microb Resour Syst. 2016;32(1):25–37.

8. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016 Jul;33(7):1870–4.

9. Garnier T, Eiglmeier K, Camus J-C, Medina N, Mansoor H, Pryor M, et al. The complete genome sequence of Mycobacterium bovis. Proc Natl Acad Sci U S A. 2003 Jun 24;100(13):7877–82.

10. Martinez AN, Talhari C, Moraes MO, Talhari S. PCR-Based Techniques for Leprosy Diagnosis: From the Laboratory to the Clinic. PLoS Negl Trop Dis. 2014 Apr 10;8(4):e2655.

11. Al AV et. *Mycobacterium lepromatosis* Lepromatous Leprosy in US Citizen Who Traveled to Disease-Endemic Areas Volume 23, Number 11—November 2017 Emerging Infectious Disease journal CDC. [cited 2017 Oct 26];

12. Han XY, Sizer KC, Tan H-H. Identification of the leprosy agent *Mycobacterium lepromatosis* in Singapore. J Drugs Dermatol JDD. 2012 Feb;11(2):168–72.

13. Jessamine PG, Desjardins M, Gillis T, Scollard D, Jamieson F, Broukhanski G, et al. Leprosy-like illness in a patient with *Mycobacterium lepromatosis* from Ontario, Canada. J Drugs Dermatol JDD. 2012 Feb;11(2):229–33.

14. Han XY, Aung FM, Choon SE, Werner B. Analysis of the Leprosy Agents *Mycobacterium leprae* and *Mycobacterium lepromatosis* in Four Countries. Am J Clin Pathol. 2014 Oct;142(4):524–32.

15. Han XY, Sizer KC, Velarde-Félix JS, Frias-Castro LO, Vargas-Ocampo F. The leprosy agents *Mycobacterium lepromatosis* and *Mycobacterium leprae* in Mexico. Int J Dermatol. 2012;51(8):952–9.

16. Vargas-Ocampo F. Diffuse leprosy of Lucio and Latapí: a histologic study. Lepr Rev. 2007 Sep;78(3):248–60.

17. Romero A, Brenes, Ibarra A, Fallas M. Clinical study of lepromatous leprosy

in Costa Rica. Int J Lepr. 1949 Jun;17(1-2):27–33.

18. Herath S, Navinan MR, Liyanage I, Rathnayaka N, Yudhishdran J, Fernando J, et al. Lucio's phenomenon, an uncommon occurrence among leprosy patients in Sri Lanka. BMC Res Notes. 2015 Nov 13;8(672).

19. Al CR et. Diffuse Multibacillary Leprosy of Lucio and Latapí with Lucio's Phenomenon, Peru Volume 23, Number 11—November 2017 Emerging Infectious Disease journal CDC. [cited 2017 Dec 27]; Available from: https://wwwnc.cdc.gov/eid/article/23/11/17-1228_article#tnF1

20. Kumari R, Thappa DM, Basu D. A fatal case of Lucio phenomenon from India. Dermatol Online J. 2008;14(2):10.

21. Ranugha PSS, Chandrashekar L, Kumari R, Thappa DM, Badhe B. Is it lucio phenomenon or necrotic erythema nodosum leprosum? Indian J Dermatol. 2013 Mar 1;58(2):160.

22. Mareen J, Madhukara J. An interesting case of Lucio phenomenon triggered by activation of hepatitis C infection. Indian Dermatol Online J. 2016 Oct;7(5):406–9.

23. Misra DP, Parida JR, Chowdhury AC, Pani KC, Kumari N, Krishnani N, et al. Lepra reaction with lucio phenomenon mimicking cutaneous vasculitis. Case Rep Immunol. 2014;2014:641989.

24. Kaur C, Thami GP, Mohan H. Lucio phenomenon and Lucio leprosy. Clin Exp Dermatol. 2005 Sep;30(5):525–7.

25. Souza CS, Roselino AM, Figueiredo F, Foss NT. Lucio's phenomenon: clinical and therapeutic aspects. Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 2000 Dec;68(4):417–25.

26. Helmer KA, Fleischfresser I, Kucharski-Esmanhoto LD, Fillus Neto J, Santamaria JR. The Lucio's phenomenon (necrotizing erythema) in pregnancy. An Bras Dermatol. 2004 Apr;79(2):205–10.

27. Monteiro R, Abreu MAMM de, Tiezzi MG, Roncada EVM, Oliveira CCM de, Ortigosa LCM. Lucio's phenomenon: another case reported in Brazil. An Bras Dermatol. 2012 Apr;87(2):296–300.

28. Choon SE, Tey KE. Lucio's phenomenon: a report of three cases seen in Johor, Malaysia. Int J Dermatol. 2009 Sep;48(9):984–8.

29. Rea TH. Latapi's lepromatosis (Lucio's leprosy) without Lucio's phenomenon in a pubertal boy. Lepr Rev. 2009 Dec;80(4):445–7.

30. Fenniche S, Benmously R, Sfia M, Daoud L, Debbiche A, Ben Ayed M, et al. [Late-occurring cutaneous vasculitis after successful treatment of diffuse lepromatous leprosy: Lucio's phenomenon]. Med Trop Rev Corps Sante Colon. 2007 Feb;67(1):65–8.

31. Golchai J, Zargari O, Maboodi A, Maboodi A, Granmayeh S. Lepromatous leprosy with extensive unusual ulcerations and cachexia. Is it the first case of Lucio's phenomenon from Iran? Int J Lepr Mycobact Dis Off Organ Int Lepr Assoc. 2004 Mar;72(1):56–9.

32. O'Reilly LM, Daborn CJ. The epidemiology of Mycobacterium bovis infections in animals and man: A review. Tuber Lung Dis. 1995 Aug 1;76(Supplement 1):1–46.

33. Nugent G, Gortazar C, Knowles G. The epidemiology of Mycobacterium bovis in wild deer and feral pigs and their roles in the establishment and spread of bovine tuberculosis in New Zealand wildlife. N Z Vet J. 2015 Jun;63 Suppl 1:54–67.

34. Pérez-Lago L, Navarro Y, García-de-Viedma D. Current knowledge and pending challenges in zoonosis caused by Mycobacterium bovis: A review. Res Vet Sci. 2014 Oct 1;97(Supplement):S94–100.

35. Baker MG, Lopez LD, Cannon MC, De Lisle GW, Collins DM. Continuing Mycobacterium bovis transmission from animals to humans in New Zealand. Epidemiol Infect. 2006 Oct;134(5):1068–73.

36. Barlow AM, Mitchell KA, Visram KH. Bovine tuberculosis in llama (Lama glama) in the UK. Vet Rec. 1999 Nov 27;145(22):639–40.

37. Leprosy – we've much left to learn, but are looking to squirrels, cows and cats for insights. J Feline Med Surg. 2017 Sep 1;19(9):977–8.

38. Pin D, Guérin-Faublée V, Garreau V, Breysse F, Dumitrescu O, Flandrois J-P, et al. Mycobacterium Species Related to *M. leprae* and *M. lepromatosis* from Cows with Bovine Nodular Thelitis. Emerg Infect Dis. 2014 Dec;20(12):2111–4.

39. O'Brien CR, Malik R, Globan M, Reppas G, McCowan C, Fyfe JA. Feline leprosy due to Candidatus "*Mycobacterium lepraefelis*": Further clinical and molecular characterisation of eight previously reported cases and an additional 30 cases. J Feline Med Surg. 2017 Sep;19(9):919–32.

40. Strain Variation in the *Mycobacterium tuberculosis* Complex: | Sebastien Gagneux | Springer [Internet]. [cited 2017 Dec 29]. Available from: //www.springer.com/gp/book/9783319643694

41. Socransky SS, Haffajee AD, Cugini MA, Smith C, Kent RL. Microbial complexes in subgingival plaque. J Clin Periodontol. 1998 Feb;25(2):134–44.

42. Mark S. Alexander the Great, Seafaring, and the Spread of Leprosy. J Hist Med Allied Sci. 2002 Jul 1;57(3):285–311.

43. Lavania M, Jadhav R, Turankar RP, Singh I, Nigam A, Sengupta U. Genotyping of *Mycobacterium leprae* strains from a region of high endemic leprosy prevalence in India. Infect Genet Evol. 2015 Dec 1;36(Supplement C):256–61.

44. Weng X, Xing Y, Liu J, Wang Y, Ning Y, Li M, et al. Molecular, ethno-spatial epidemiology of leprosy in China: novel insights for tracing leprosy in endemic and non endemic provinces. Infect Genet Evol J Mol Epidemiol Evol Genet Infect Dis. 2013 Mar;14:361–8.

45. Yuan Y, Wen Y, You Y, Xing Y, Li H, Weng X, et al. Characterization of *Mycobacterium leprae* Genotypes in China—Identification of a New Polymorphism C251T in the 16S rRNA Gene. PLOS ONE. 2015 Jul 21;10(7):e0133268.

46. WHO. Global leprosy update, 2016: accelerating reduction of disease burden. 2017 p. 501–20. (35). Report No.: 92.

47. Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the *Mycobacterium tuberculosis* Beijing genotype strains. Lancet Infect Dis. 2010 Feb 1;10(2):103–11.

48. Cardona-Castro N, Cortés E, Beltrán C, Romero M, Badel-Mogollón JE, Bedoya G. Human Genetic Ancestral Composition Correlates with the Origin of *Mycobacterium leprae* Strains in a Leprosy Endemic Population. PLoS Negl Trop Dis. 2015 Sep 11;9(9):e0004045.

49. Manabe YC, Dannenberg AM, Tyagi SK, Hatem CL, Yoder M, Woolwine SC, et al. Different Strains of *Mycobacterium tuberculosis* Cause

Various Spectrums of Disease in the Rabbit Model of Tuberculosis. Infect Immun. 2003 Oct;71(10):6004–11.

50. Kato-Maeda M, Shanley CA, Ackart D, Jarlsberg LG, Shang S, Obregon-Henao A, et al. Beijing Sublineages of *Mycobacterium tuberculosis* Differ in Pathogenicity in the Guinea Pig. Clin Vaccine Immunol CVI. 2012 Aug;19(8):1227–37.

51. Ribeiro SCM, Gomes LL, Amaral EP, Andrade MRM, Almeida FM, Rezende AL, et al. *Mycobacterium tuberculosis* Strains of the Modern Sublineage of the Beijing Family Are More Likely To Display Increased Virulence than Strains of the Ancient Sublineage. J Clin Microbiol. 2014 Jul;52(7):2615–24.

52. Coscolla M, Gagneux S. Does *M. tuberculosis* genomic diversity explain disease diversity? Drug Discov Today Dis Mech. 2010;7(1):e43–59.

53. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug resistant tuberculosis. Nat Genet. 2013 Jul;45(7):784–90.

54. Daniels CC, Rogers PD, Shelton CM. A Review of Pneumococcal Vaccines: Current Polysaccharide Vaccine Recommendations and Future Protein Antigens. J Pediatr Pharmacol Ther JPPT. 2016;21(1):27–35.

55. Robinson WH. Sequencing the functional antibody repertoire—diagnostic and therapeutic discovery. Nat Rev Rheumatol. 2015 Mar;11(3):171–82.

56. WHO. World Health Organization. Guidelines for the programmatic management of drug-resistant tuberculosis: 2011 update. . 2013. Report No.: WHO/HTM/TB/2011.6.

57. Gillis TP, Williams DL. Dapsone resistance in *Mycobacterium leprae*. Lepr Rev. 2000 Dec;71 Suppl:S91–5.

58. Kohanski MA, Dwyer DJ, Collins JJ. How antibiotics kill bacteria: from targets to networks. Nat Rev Microbiol. 2010 Jun;8(6):423–35.

59. Nakata N, Kai M, Makino M. Mutation Analysis of the *Mycobacterium leprae* folP1 Gene and Dapsone Resistance. Antimicrob Agents Chemother. 2011 Feb;55(2):762–6.

60. Van Kessel JC, Hatfull GF. Recombineering in *Mycobacterium tuberculosis*. Nat Methods. 2007 Feb;4(2):147–52.

61. Van Kessel JC, Hatfull GF. Mycobacterial recombineering. Methods Mol Biol Clifton NJ. 2008;435:203–15.

62. Tufariello JM, Kerantzas CA, Vilchèze C, Calder RB, Nordberg EK, Fischer JA, et al. The Complete Genome Sequence of the Emerging Pathogen *Mycobacterium haemophilum* Explains Its Unique Culture Requirements. mBio. 2015 Dec 31;6(6):e01313–5.

Curriculum vitae

Charlotte Avanzi

Grand rue 24
1180 Rolle
Switzerland

French, age 30

Charlotte.avanzi@gmail.com
+41 79 615 87 40
https://orcid.org/0000-0002-1062-4058
www.linkedin.com/in/charlotte-avanzi
@CharlotteAvanzi

Education

2006-2014	**PharmD (Industry section), Université Grenoble Alpes**
2012-2013	**Master of Science (MSc) program:** Health Engineering and Medicine specialty Industry *in vitro* Diagnosis, Université Grenoble Alpes Supervisor: Dr. Jean Breton (jean.breton@ujf-grenoble.fr)

Professional experience & trainings

2013-present	**PhD in Molecular Life Science** [1–7] **/ teaching assistant** Supervisor: Prof. Stewart Cole (stewart.cole@epfl.ch) Ecole Polytechnique fédérale de Lausanne, Switzerland
2013	**6 months training at Fondation Mérieux, Lyon, France** Development and validation of a new detection tool for *S. pneumoniae* with the Luminex xTAG technology
2012	**3 months training at Hong Kong University pharmacology department / Queen Mary Hospital of Hong Kong** Development of a dosage method to demonstrate the interaction between cyclosporine and Chinese medicinal plants (HPLC method) and interactions between traditional Chinese medicine and cyclosporin through CYP3A4 inhibition (enzymatic assays). [8] Supervisor: Dr. Sophie Dufaÿ
2012	**3 months training at Centre Hospitalier Universitaire as pharmacist trainee, Grenoble, France** Clinical pharmacy support in psychiatry unit and involvement in a psychiatry clinical assay supervision
2011	**3 months training at Institut Pasteur de Madagascar, Antananarivo, Madagascar** Cysticercosis: study of the immune response of human body against the parasite cyst fluid (ELISA, Western blot, flow cytometry) Supervisor: Dr. Ronan Jambou
2008-2013	Seasonal work at SDIS 73, Employee in a Pharmacy, 1st year student teaching in cellular biology (Grenoble)

Skills

Scientific

- Molecular biology (DNA, RNA extraction, PCR)
- DNA seq: preparation and analysis
- Bacteriology: molecular cloning and mutant generation
- BSL-2 and BSL-3 laboratory work
- Module 1 animal experimentation (RESAL)

IT

- Data analysis and statistics in R
- NGS analysis through a pipeline
- Basics in Bach and R (data analysis and statistics)
- Data analysis in Prism7

Communication, teaching and management

- Public presentation in French and English
- Teaching to collaborators (five trainees: from 2 weeks to 3 months supervision)
- Team working and project management
- Report and publication writing in French and English

Languages:

- French: native
- English: fluent
- Italian: beginner

Conferences & memberships

Conferences

- ECCMID 2015 – Copenhagen, Denmark; (ePoster presentation)
- IMMEM XI 2016 – Estoril, Portugal; (oral presentation)
- ILC 2016 – Beijing, China; (oral presentation)
- Collège de France, 2016 – Paris, france (Invited speaker: https://www.college-de-france.fr/site/philippe-sansonetti/seminar-2016-12-07-17h30.htm)
- ASM microbe 2017: New Orleans, USA (poster presentation)
- Fondation Raoul Follereau annual congress 2017 – Paray-le-Monial, France (oral presentation)
- Colloque INSERM on applications and perspectives for whole genome sequencing for mycobacteria, 2017 – Paris, France (invited speaker)
- RICAI 2017 – Paris, France (invited speaker)

Memberships

- Swiss Society for Molecular and Cellular Biosciences (SSMCB)/ Life Sciences Switzerland (LS2) since February 2014
- The European Society of Clinical Microbiology and Infectious Diseases (ESCMID) since April 2015
- The American Society for Microbiology (ASM) since December 2016

Fellows

- EMBO short term fellowship – 15th June 2015 to 4th of July 2015– Visit to the EberhardKarls – Universitât Tübingen

Courses

- Institut Pasteur course in immunology (2 weeks, Antananarivo, Madagascar, 2011)
- Computer skills for biological research (3 days, CUSO, Geneva, 2016)
- Summer course in Immunology (1 week, Université de Lausanne, 2015)

Publications

1. 1. P. Singh. A. Benjak. S. Carat. M. Kai. P. Busso. **C. Avanzi**. A. Paniz-Mondolfi. C. Peter. K. Harshman. J. Rougemont. . M. Matsuoka. S.T. Cole Genome-wide re-sequencing of multi-drug resistant *Mycobacterium leprae* Airaku-3*Clin Microbiol Infect* 2014. 20(10):619-22

2. P. Singh. A. Benjak. V.J. Schuenemann. A. Herbig. **C. Avanzi**. P. Busso. K. Nieselt. J. Krause. L. Vera-Cabrera. S.T. Cole Insight into the evolution and origin of leprosy bacilli from the genome sequence of *Mycobacterium lepromatosis – PNAS* 2015. 112(14): 4459-4464

3. L. Vera-Cabrera. W. Escalante-Fuentes. S. Ocampo-Garza. J. Ocampo-Candiani. C. Molina-Torres. **C. Avanzi**. A. Benjak. P. Busso. P. Singh. and S. T. Cole *Mycobacterium lepromatosis* infections in Nuevo León. México – *J Clin Microbio* March 2015

4. Transmission of drug-resistant leprosy in Guinea-Conakry detected by molecular epidemiological approaches – **C. Avanzi**, P. Busso, A. Benjak, C. Loiseau, A. Fomba, G. Doumbiab, I. Camara, A. Lamou, G. Sock, T. Drame, M. Kodio, F. Sakho, S.O. Sow, S. T. Cole, R.C. Johnson *Clin Infec Dis.* 2016; 1;63(11):1482-1484

5. Red squirrels in the British Isles are infected with leprosy bacilli – **C. Avanzi***, J. del-Pozo*, A. Benjak*, K. Stevenson, V. R. Simpson, P. Busso, J.McLuckie, C. Loiseau, C. Lawton, J. Schoening, D. J. Shaw, J. Piton, L. Vera-Cabrera, J. S. Velarde-Felix, F. McDermott, S. V. Gordon, S. T. Cole, A. L. Meredith – *Science*. 2016; 354(6313): 744-747

6. Whole genome sequencing distinguishes between relapse and reinfection in recurrent leprosy cases M. M. A. Stefani*, **C. Avanzi***, S. Bührer-Sékula, A. Benjak, C. Loiseau, P. Singh, M. A. A. Pontes, H. S. Gonçalves, E. M. Hungria, P. Busso, J. Piton, M. I. S. Silveira, R. Cruz, A. Schetinni, M. B. Costa, M. C. L. Virmond, S. M. Diorio, I. M. F. Dias-Baptista, P. S. Rosa, M. Matsuoka, M. L. F. Penna, S. T. Cole, G. O. Penna – *Plos Negl. Trop. Dis*. 2017; 11(6)

7. A. Benjak, TP. Honap, **C. Avanzi**, E. Becerril-Villanueva, I. Estrada-Garcia, O. Rojas-Espinosa, AC. Stone, ST Cole Insights from the Genome Sequence of *Mycobacterium lepraemurium*: Massive Gene Decay and Reductive Evolution. *mBio* 2017;8(5):e01283–17.

8. S. Dufaÿ. A.Worsley. A. Monteillier. **C. Avanzi**. J. Sy. T.F. Ng. J.M. Garcia. M.F. Lam. P. Vanhoutte. I.C. K. Wong Herbal tea extracts inhibit Cytochrome P450 3A4 in vitro – *J Pharm Pharmacol* 2014. 66(10):1478-90

Extra-scientific activities

- Sapeur Pompier volontaire, SDIS73, France (2004-2015)
    - Sapeur 1ère classe au Centre de Secours principal de St Jean de Maurienne (2004-2013)
    - Pharmacien Capitaine au Service de Santé et de Secours Médical de Chambéry (2014-2015) – missions au SDMIS; groupe de travail sur le risqué biologique chez les sapeurs pompiers
- Hiking, skiing, ski touring, climbing & bouldering, alpinism, member of le Club Alpin Suisse since 2016 (section Carougeoise)
- Diving (PADI Open Water Diver)
- Travelling and enjoying time with my friends and family

Remerciements

A la leprosy team aka la team squirrel

Cette thèse n'aurait jamais pu être possible sans la contribution des membres de la leprosy team aka:
Le léprologue de langue française Mr Pippo, mon binôme, merci de m'avoir appris la biologie moléculaire et comment ranger les tubes dans la centrifugeuse, ça me servira surement toute ma vie. Merci de m'avoir supporté pendant ces 4 ans et pour toute ton aide dans les projets leper et autres. Merci pour les wiki-cours de Suisse et autres les quizzs de culture general, les corrections des fautes d'orthographe, pour les imitations des différents accents et d'être venu chasser l'écureuil roux avec moi.

Dédé, merci pour toute l'aide avec le séquençionnage, les corrections de papier, de rapports, les analyses de dernières minutes et tout le temps que tu as passé à m'expliquer les petits trucs de inkscape et autres softwares. Merci pour toute ta patience, je n'aurais jamais pu mener tous ces projets sans ton aide, j'ai beaucoup appris avec toi. Merci pour l'organisation des hikes, tes blagues et de m'avoir accepté dans ton bureau. Ca a été des super années!

A mon petit génie Chloé, merci beaucoup pour ta bonne humeur et toute ton aide avec cette matière non intuitive qu'est la bioinformatique. Merci pour la pipeline, ta motivation et tes blagues de geek. 6 mois c'était beaucoup trop court mais ça a été un très grand plaisir de t'avoir un peu pour moi tout ce temps.

Thanks a lot to our wold leprosy expert Prof. Singh for his guidance and help in my first steps in the leprosy field.

The last but not the least; Stewart, I really would like to thank you. Not only supervision but I would also like to particularly thank you for introducing me into the leprosy world, for your guidance and advises at all steps, your trust and the life teaching. I learned a lot from you and I evolved in a very motivated environment, it was really a pleasure and thank you for that.

Au banana power club
Raphayel, merci pour les imitations, pour le show et les longues heures de baratin.
Palomita, merci de m'avoir enseigné tous ces jolis mots en espagnol!
Jérémito, merci pour l'écureuil en peluche et tes blagues de français
Fooch, mon petit "boulet", merci pour ton rire contagieux et ta constante bonne humeur!
LupLup, merci pour tous, au labo et mais aussi merci pour les sorties à Europa Park, les soirées chez Gina, les apéros imprévus et ton amitié!

Merci à tous les UPCOOlers
Stef and Antho thank you so much for all your help during these years and thanks for supporting me with all the lab stuffs and ordering.
Nina: danke schoen for your help during these four years, thanks for the organization of all the events in the lab!

Merci aux anciens UPCOlers pour toute leur aide dans la réalisations de ces projets!
Cécile P, Suzanne D., Trent, Joao, Gaëlle, Ruben, Benoît, Benjamin et Ye

Aux UPBLO:
Merci à Charles pour sa positivité contagieuse, ses conseils, les bières belges et pour le tandem!
Merci à Noémie pour le super dessin!
Merci à Sandrine pour son aide avec les extractions d'ADN
Merci à Ivan pour les statistiques
Merci à Candice parce que c'est Candice ;-)
Danke schoen to my mentor Mélanie for all advises and nice discussions during my thesis!

Aux UPLEM
Merci Sam pour ta bonne humeur transitoire et de m'avoir emmené sur les Perrons de Vallorcine! Muchas gracias guapo!
Fanny, merci beaucoup pour toute l'aide avec la qPCR! Et le tandem ;-) qui inclut le tour en bateau!
Merci à Olivier B. pour sa disponibilité et son écoute pendant ces deux dernières années.

Thank to all member of my jury for having accepted to judge this work

A mes collaborateurs

A mes soutiens sans qui cette these n'aurait jamais pu avoir lieu
Dr J. Breton, Mélina Messaoudi, Glaucia paranhos-Baccala

Au Prof. Renée Grillot que j'admire et qui m'a donné l'envie de faire de la recherche

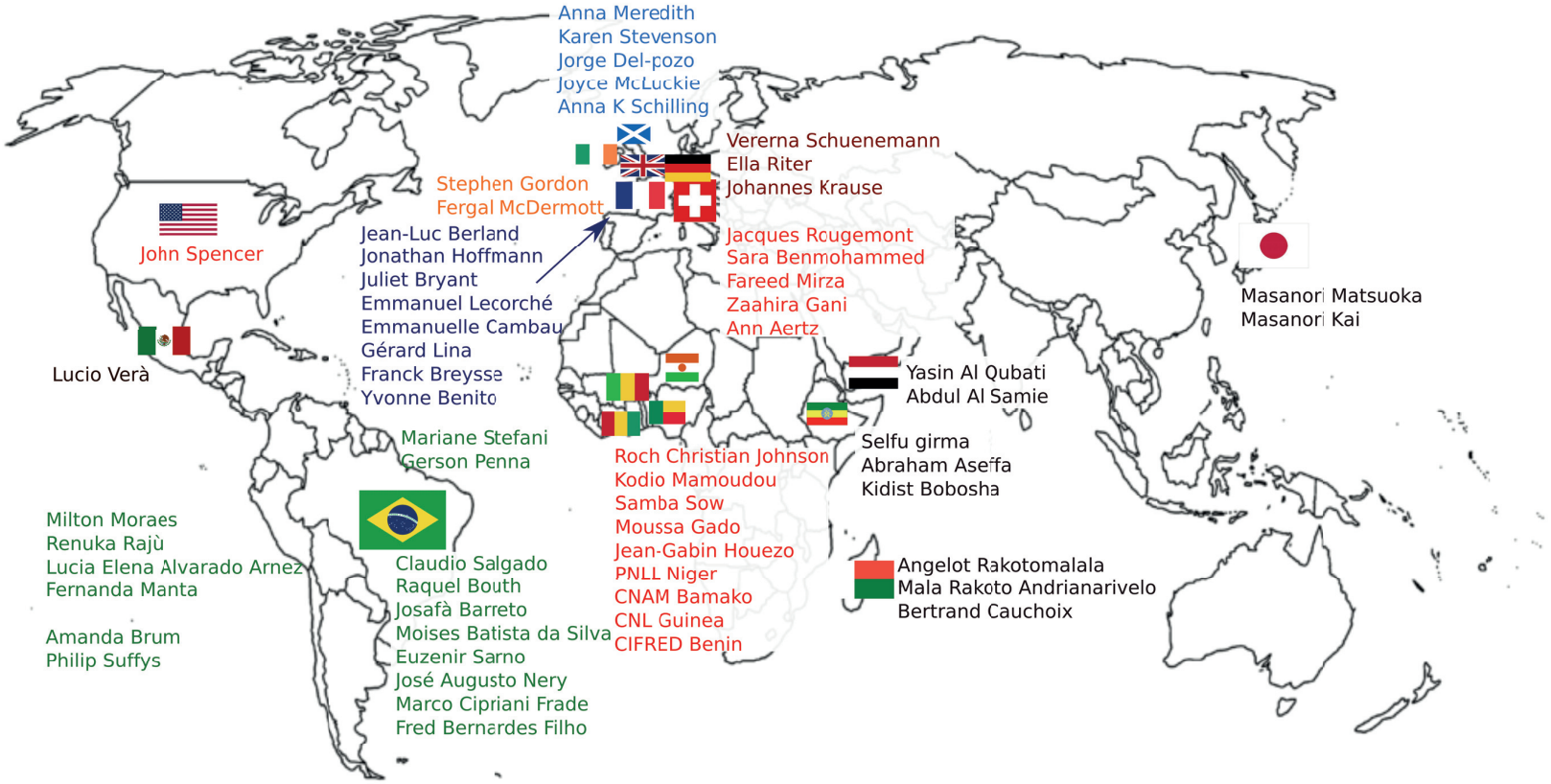
Aux copains pour leur soutien quotidien, leur petits mots doux quand le moral n'est pas au beau fixe et pour les supers week ends improvisés: Lisa, Bibar, doriane, Laurette, Heidi, Fadia, Gui, Jérém, Mouton, Laurent, Tristan, Gauthier, Thomas, Rominou, Damien & Marie, Marie et pépère, Margueu, Nico, Glouglou, Nonal, la Beast et Sushi, Clothilde et Bébesse, Théo & Joëlle, Léoninie

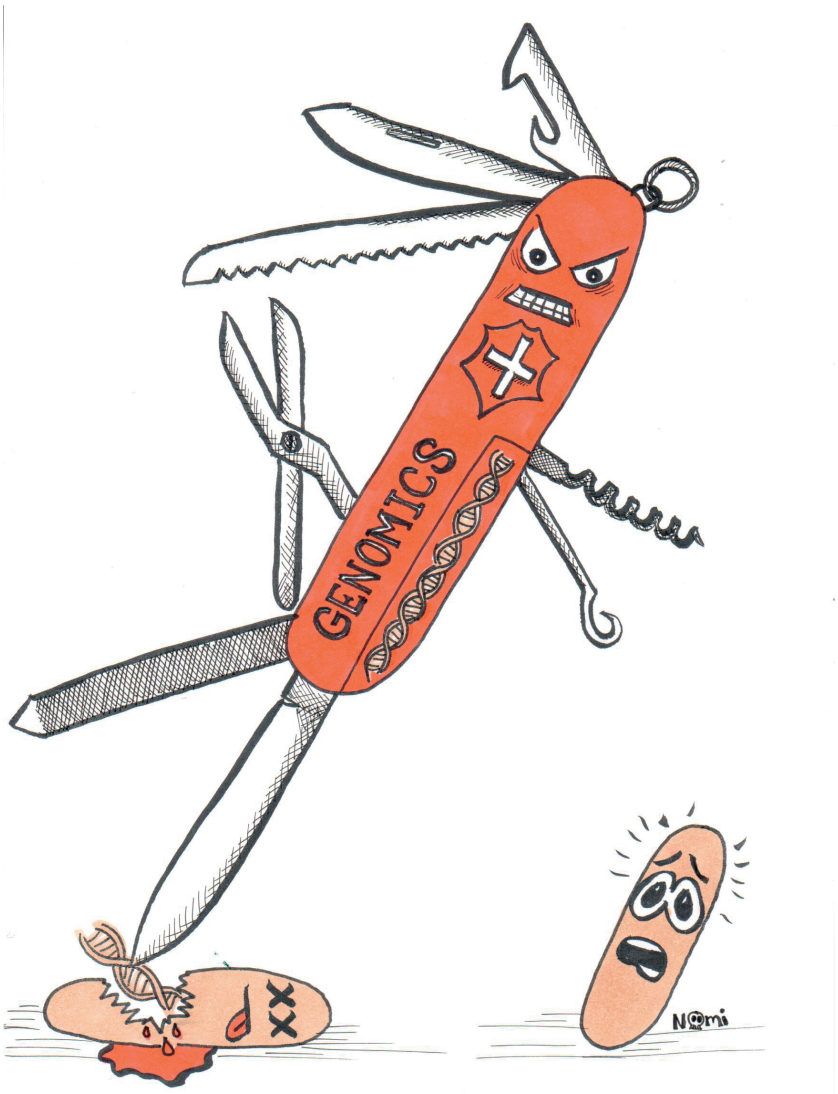
Aux copains du CAS vin rouge qui m'ont permis de prendre de grands bols d'air pendant cette thèse
Andréa, Aurélie, Ariane, Michel, Laurent et Etienne

Aux copains de Totem, ma seconde échapattoire

A la mairie de Pontamafrey-Montpascal & les employés pour leur acceuil durant la rédaction

A ma famille,
Merci à mes parents pour toute leur aide durant cette these et bien avant. Vous êtes mes exemples et ma plus grande source d'inspiration.
A mes frères, merci pour votre soutien sans failles, vos messages et de veiller toujours sur moi.
A mes couz' Nina, Cam et Marianne: merci d'être venue me voir, me chercher et me soutenir pendant les moments difficiles.
A tous les autres, Nadine, Philippe, Patricia, J-P, Mémé, Anne Marie, Marie Grâce, Jean-Yves, Andréa, Granini & les Audras/Monteilliers qui ont un jour eu un geste, un mot de soutien ou même montrer de l'intérêt pour mon sujet juste pour me faire plaisir. ☺ Vous êtes géniaux!
A Brigitte et Robert et toute ma belle famille, merci de vous être occupé de moi pendant toutes ces années
Et enfin à Aymeric,
S'il y a quelqu'un à qui je dois dédier cette thèse c'est bien à toi. Merci d'avoir été tantôt Aymeric Robuchon, Aymeric Steck, Aymeric Poppins, Aymeric Freud et ma patate. Merci pour ta compréhension, tes conseils et de me pousser tout le temps toujours plus loin. Sans toi tout aurait été beaucoup plus difficile. Tu es ma plus grande motivation.


Anna Meredith
Karen Stevenson
Jorge Del-pozo
Joyce McLuckie
Anna K Schilling
Vererna Schuenemann
Ella Riter
Johannes Krause
Stephen Gordon
Fergal McDermott
John Spencer
Jean-Luc Berland
Jonathan Hoffmann
Juliet Bryant
Emmanuel Lecorché
Emmanuelle Cambau
Gérard Lina
Franck Breysse
Yvonne Benito
Jacques Rougemont
Sara Benmohammed
Fareed Mirza
Zaahira Gani
Ann Aertz
Masanori Matsuoka
Masanori Kai
Lucio Verà
Yasin Al Qubati
Abdul Al Samie
Mariane Stefani
Gerson Penna
Roch Christian Johnson
Kodio Mamoudou
Samba Sow
Moussa Gado
Jean-Gabin Houezo
PNLL Niger
CNAM Bamako
CNL Guinea
CIFRED Benin
Selfu girma
Abraham Aseffa
Kidist Bobosha
Milton Moraes
Renuka Rajù
Lucia Elena Alvarado Arnez
Fernanda Manta
Amanda Brum
Philip Suffys
Claudio Salgado
Raquel Bouth
Josafà Barreto
Moises Batista da Silva
Euzenir Sarno
José Augusto Nery
Marco Cipriani Frade
Fred Bernardes Filho
Angelot Rakotomalala
Mala Rakoto Andrianarivelo
Bertrand Cauchoix



©Noémie Mattey